Theoretical Computer Science 71 (1990) 281-346
North-Holland

FUNDAMENTAL STUDY

SOME DECISION PROBLEMS ABOUT CONTROLLED REWRITING SYSTEMS*

G. SÉNIZERGUES**
U.A. CNRS No. 276, Laboratory of Computer Science, Bordeaux 1 University, 351 Cours de la Libération, 33405, Talence Cedex, France

Communicated by M. Nivat
Received January 1988
Revised July 1988

Abstract. A description of deterministic context-free languages by some kind of finitely generated right-congruences (the sets of relations generating such right-congruences are called controlled rewriting systems or c-systems) has been given previously. In the first part of this paper, we study several decision problems on c-systems, namely the confluence problem, the equivalence problem, the refinement problem, the class equivalence problem and the class inclusion problem. In the second part of the paper, we deduce, from the main theorem given previously, a characterisation of dcfls by means of finitely generated congruences (the sets of relations generating such congruences are now ordinary finite semi-Thue systems). We study three decision problems on finite semi-Thue systems, namely the class equivalence problem, the word problem for the syntactic congruence of one class and the partial confluence problem. All the problems are investigated in several classes of c-systems or semi-Thue systems. For every class we give an answer which may be "yes" (the problem is decidable), "no" (the problem is undecidable) or "eq" which means that the problem is recursively equivalent to the equivalence problem for dpda.

Contents

1. Introduction 282
2. Preliminaries 283
 2.1. Deterministic pushdown automata 283
 2.2. Finite automata 284
 2.3. Congruences 285
 2.4. Controlled rewriting systems 285
3. Some properties of controlled rewriting systems 289
 3.1. Deterministic context-free languages and c-systems 289
 3.2. Rational and finite c-systems 289
 3.3. Leftmost reduction generated by a c-system 294
 3.4. Total reduction generated by a confluent, strict c-system 295

* This work has been supported by the "Programme de Recherches Coordonnées en Mathematiques et Informatique".
** This work has been realised by the author when he was researcher at LITP (U.A. au CNRS No. 248).

4. Some decision problems about strict finite c-systems 296
4.1. Summary of the problems 296
4.2. Proofs of the results 298
5. Decision problems about finite semi-Thue systems in connection with the equivalence problem for dpda 317
5.1. A representation of dcfls by left-basic, confluent, finite semi-Thue systems 318
5.2. Some decision problems about finite semi-Thue systems 329
6. Conclusion 343
References 345

1. Introduction

The notion of a controlled rewriting system (c-system) was introduced in [12] and [14] in connection with formal language theory. In [36] the results of [15] were improved in the following way: it is established that the deterministic cfls are exactly those languages which can be defined as $[R]_{\overset{*}{\leftrightarrow}_S}$ where R is a rational language and $\overset{*}{\leftrightarrow}_S$ is the right-congruence generated by a basic, confluent, strictly length-reducing, finite c-system S.

A finite c-system over an alphabet X is a part S of $X^* \times X^* \times X^*$ which can be decomposed as a finite union: $S = \bigcup_{i=1}^{n} R_i \times \{u_i\} \times \{v_i\}$ where the R_i are rational sets and u_i, v_i are words. Such a c-system defines a reduction $\vdash_S : f \vdash_S g$ iff $f = r_i v_i s$, $g = r_i u_i s$ for some $i \in [1, n]$, $r_i \in R_i$ and $s \in X^*$. The equivalence relation $(\vdash_S \cup \vdash_S^{-1})^*$, noted $\overset{*}{\leftrightarrow}_S$, is the right-congruence generated by S. This type of right-congruence generalises the congruences generated by a finite semi-Thue system (this is clear) or by a rational semi-Thue system (this is shown here). The restriction "basic" mentioned above is a combinatorial property which is defined in [25] for semi-Thue systems and generalised in [12] to c-systems.

First we investigate for c-systems some decision problems which have already been investigated for finite [25, 26], rational [27, 28, 24] or context-free [10] semi-Thue systems. These problems are the confluence problem, the equivalence problem, the refinement problem, the class equivalence problem and the class inclusion problem (see in Section 4.1 a detailed description of these problems and the results obtained). As far as we know, c-systems have similar decidability properties as rational semi-Thue systems (an exception is the positive result of [24] cited in Section 6, while we give the Proposition 4.5 a negative result concerning the same problem for c-systems instead of rational semi-Thue systems).

As the equality between one language of the form $[R]_{\overset{*}{\leftrightarrow}_S}$ (where R is rational and S is a basic, confluent, length-reducing, rational semi-Thue system) and one dcfl is decidable [35, Part IV], the similarity between rational semi-Thue systems and finite c-systems (with respect to decision problems) strengthens the hope that the equality between two dcfls is also decidable. It is also noteworthy that the equality (and even the inclusion) between the right-congruences generated by two basic, confluent, length-reducing, finite c-systems is decidable (Proposition 4.2) while the equivalent problem for dcfls reduces (by Theorem 3.1 of [36]) to $[f]_{\overset{*}{\leftrightarrow}_{S_1}} = [f]_{\overset{*}{\leftrightarrow}_{S_2}}$? where f is a word and S_1, S_2 are two basic, confluent, length-reducing, finite c-systems (this last problem is an instance of what we called the "class equivalence problem" for basic, confluent, length-reducing, finite c-systems).

In the second part we deal with the classical notion of a finite semi-Thue system. Our starting point is Theorem 5.12 which gives some new links between dcfls and finite semi-Thue systems. In [32] is defined the notion of a "left-basic" semi-Thue system which generalises the notion of a "basic" semi-Thue system and remains sufficient to ensure context-freeness of the classes: every class for a congruence generated by a left-basic, length-decreasing, confluent, finite semi-Thue system is a dcfl (Theorem 7.6 of [32]). Theorem 5.12 gives a characterisation of dcfls by means of congruences generated by left-basic, length-decreasing, confluent, finite semi-Thue systems. This result leads to the recursive equivalence of the general equivalence problem for dcfls with any one of the three following problems (Theorem 5.17):

(1) the equivalence problem for a strict subfamily of the family of dcfls: namely the family of languages of the form $[f]_{\leftrightarrow_S}$ for some word f and some left-basic, length-decreasing, confluent, finite semi-Thue system S;

(2) the word problem for the syntactic congruence of a language in the subfamily described above;

(3) the problem of whether $\langle f\rangle_{\to_S} = [f]_{\leftrightarrow_S}$ where S is a left-basic, length-decreasing, finite semi-Thue system (here $\langle f\rangle_{\to_S}$ denotes $\{g \in X^*,\ g \vdash^*_S f\}$, i.e. the set of ancestors of f); we call this last problem the partial confluence problem.

2. Preliminaries

Here we describe the notation, give some technical definitions and some basic results about deterministic pushdown automata (Section 2.1), finite automata (Section 2.2), congruences (Section 2.3) and controlled rewriting systems (Section 2.4).

2.1. Deterministic pushdown automata

A deterministic pushdown automaton is a sextuple $A = \langle X, Y, Q, d, q_0, y_0\rangle$ where X, Y, Q are three finite sets called, respectively, the *input-alphabet*, the *stack-alphabet* and the *set of states*. d, the transition function, is a partial mapping: $YQ \times (X \cup \{\varepsilon\}) \to Y^*Q$. d is such that, for every $yq \in YQ$, if there exists $x \in X$ such that $d(yq, x)$ is defined then $d(yq, \varepsilon)$ is undefined. q_0 is the *initial state* of A, $y_0 \in Y$ is the initial stack-symbol of A. We call *configuration* of A, every word $c = wq$ where $w \in Y^*$ and $q \in Q$. We think of w as the stack content written with its topmost symbol on the right. A configuration $c = q \in Q$ corresponds to a vacuous stack-content and a state q. We call *mode* every element $zq \in (Y \cup \{\varepsilon\})Q$. The mode of configuration wq is defined by

- if $w = \varepsilon$, then mode $(wq) = q$,
- if $w = w'z$, where $w' \in Y^*$ and $z \in Y$, then mode $(wq) = zq$.

We call *ε-mode* every mode $yq \in YQ$ such that $d(yq, \varepsilon)$ is defined. A *reading-mode* is a mode $yq \in YQ$ such that $d(yq, x)$ is defined for at least one letter $x \in X$. We call *ε-free mode* every mode $zq \in (Y \cup \{\varepsilon\})Q$ which is not an ε-mode (i.e. every zq such that either $z = \varepsilon$ or zq is a reading-mode or for every $x \in X \cup \{\varepsilon\}$, $d(zq, x)$ is undefined). Let $q, q' \in Q$, $w, w' \in Y^*$, $y \in Y$, $f \in X^*$ and $a \in X \cup \{\varepsilon\}$; we note $(wyq, af) \vdash_A (ww'q', f)$ if $d(yq, a) = w'q'$. $\vdash^*_A$ is the reflexive and transitive closure of $\vdash_A$. For every wq, $w'q \in Y^*Q$ and $f \in X^*$, we note $wq \xrightarrow{f}_A w'q'$ iff $(wq, f) \vdash^*_A (w'q', \varepsilon)$. The underscript A will be omitted when no confusion is possible. We note $wq \Vdash^f_A w'q'$ iff $wq \xrightarrow{f}_A w'q'$ and the mode of $w'q'$ is ε-free (in other words, iff A, starting with configuration wq, can read f without reaching a loop and stops its longest calculus (reading this fixed word f) in configuration $w'q'$).

Given a set F of modes, the language accepted by A with set of final modes F is

$$L(A, F) = \{f \in X^* \mid \exists c \in Y^*Q, y_0 q_0 \underset{A}{\xrightarrow{f}} c \text{ and mode } (c) \in F\}.$$

A language $L \subset X^*$ is said to be *deterministic* (resp. *strict deterministic*) *context-free* iff there exists a dpda A and a set of final modes $F \subset (Y \cup \{\varepsilon\})Q$ (resp. $F \subset Q$) such that $L = L(A, F)$.

We recall that the equivalence problem for dpda consists of determining, given two dpda A_1, A_2 and sets of final modes F_1, F_2, whether the languages $L(A_1, F_1)$ and $L(A_2, F_2)$ are equal. Though this problem has been shown to be decidable for some subclasses of dpda [38, 31, 35], it is still unknown whether the general problem is decidable or not.

2.2. Finite automata

A *finite automaton* is a 5-tuple $B = \langle X, Q, q_-, d, Q_+ \rangle$ where X, Q are finite sets called, respectively, the *input-alphabet* and the *set of states*, Q_+, the set of *terminal states* is a subset of Q, $q_- \in Q$ is the *initial state* of B, and d, the *transition function*, is a mapping

$$d : Q \times (X \cup \{\varepsilon\}) \to \mathcal{P}(Q).$$

B is said to be *deterministic* iff

(1) $\forall q \in Q$, $d(q, \varepsilon) = \emptyset$;

(2) $\forall q \in Q$, $\forall x \in X$, $|d(q, x)| \leq 1$.

A deterministic finite automaton B is said to be *complete* iff

(3) $\forall q \in Q$, $\forall x \in X$, $|d(q, x)| = 1$.

We shall call *calculus* of B, every finite sequence $C = q_0, x_1, q_1, \ldots, x_i, q_i, \ldots, x_n, q_n$ such that $n \in \mathbb{N}$; for every $i \in [1, n]$, $x_i \in X \cup \{\varepsilon\}$; for every $i \in [0, n]$, $q_i \in Q$; and for every $i \in [0, n-1]$, $q_{i+1} \in d(q_i, x_{i+1})$. If $f = x_1 x_2 \ldots x_i \ldots x_n$, we note $q_0 \to^f q_n$. In the case where B is deterministic, we use the notation $q_0 . f = q_n$.

Given a subset Q' of Q, the language recognised by B with Q' as set of terminal states, noted $L(B, Q')$, is defined by

$$L(B, Q') = \{f \in X^* \mid \exists q' \in Q', q_0 \xrightarrow{f} q'\}.$$

We use the notation $L(B)$ for $L(B, Q_+)$, $L(B)$ is the *language recognised by B*. We call regular (or rational) language over X, every language L over X such that there exists a finite automaton B recognising L.

2.3. *Congruences*

Let $(M,.)$ be a monoid and let $\mathcal{R}$ be an equivalence relation on M. We consider the following properties:

$$\forall x \in M, \forall y \in M, \forall z \in M, \quad x \mathcal{R} y \Rightarrow x.z \mathcal{R} y.z \tag{1}$$

$$\forall x \in M, \forall y \in M, \forall z \in M, \quad x \mathcal{R} y \Rightarrow z.x \mathcal{R} z.y. \tag{2}$$

We say that $\mathcal{R}$ is a right-congruence (resp. left-congruence) iff $\mathcal{R}$ fulfils property (1) (resp. (2)). We say that $\mathcal{R}$ is a congruence iff $\mathcal{R}$ is both a right-congruence and a left-congruence.

2.4. *Controlled rewriting systems*

2.4.1. *General properties of reductions*

Let E be a set and let $\vdash$ be a binary relation which we call *direct reduction*. We shall follow the framework of [21]. Hence we shall consider the binary relations $\vdash^i$ (for every $i \geq 0$), $\overset{+}{\vdash}$, $\overset{*}{\vdash}$. The relation $\vdash^{-1}$ will be denoted by $\rightarrow$ (or sometimes by $\dashv$). We call it the *direct derivation*.

The relation $\vdash \cup \dashv^{-1}$ is denoted by $\leftrightarrow$. The three relations $\overset{*}{\vdash}$, $\overset{*}{\rightarrow}$, $\overset{*}{\leftrightarrow}$ are called, respectively, the *reduction*, the *derivation* and the *equivalence* generated by $\vdash$.

About $\vdash$ we consider the properties of being *confluent*, *locally confluent*, *Church-Rosser*, *noetherian*, as they are defined in [21].

An element $e \in E$ is said to be $\vdash$*-irreducible* (or irreducible modulo $\vdash$) iff there is no e' such that $e \vdash e'$. We denote by $\mathrm{Irr}(\vdash)$ the set of all elements which are irreducible modulo $\vdash$. e is said to be $\vdash$*-reducible* (or reducible modulo $\vdash$) iff there exists some e' such that $e \vdash e'$.

Let A be a subset of E. We use the following notation:

$$\langle A \rangle_{\overset{*}{\rightarrow}} = \{e \in E \mid \exists a \in A, a \overset{*}{\rightarrow} e\},$$

$$[A]_{\overset{*}{\leftrightarrow}} = \{e \in E \mid \exists a \in A, a \overset{*}{\leftrightarrow} e\}.$$

We recall the classical lemma.

2.1. Lemma. *Let $\vdash$ be a noetherian reduction on E. $\vdash$ is confluent iff for every $e \in E$, $|[e]_{\overset{*}{\leftrightarrow}} \cap \mathrm{Irr}(\vdash)| = 1$.*

Let us now consider the particular case where $E = X^*$, the free monoid generated by some alphabet X. $\vdash$ is said to be l*-reducing* (resp. *strictly* l*-reducing*) iff it reduces (resp. reduces strictly) the length, that is,

$$\forall f \in X^*, \forall g \in X^*, \quad f \vdash g \Rightarrow |f| \geq |g| \quad (\text{resp. } |f| > |g|).$$

In the following we use the abbrevation "strict" for "strictly l-reducing". A *valuation* over X^* is an homomorphism $\nu : (X^*,.) \to (\mathbb{N}, +)$ such that, for every $x \in X$, $\nu(x) \neq 0$. $\vdash$ is said to be *v-reducing* (resp. strictly v-reducing) iff there exists a valuation ν over X^* such that $\vdash$ reduces (resp. strictly reduces) the valuation of the words, that is,

$$\forall f \in X^*, \forall g \in X^*, \quad f \vdash g \Rightarrow \nu(f) \geqslant \nu(g) \quad (\text{resp. } \nu(f) > \nu(g)).$$

In the following we use the abbrevation "v-strict" for "strictly v-reducing".

2.4.2. *Controlled rewriting systems*

The notion of a controlled rewriting system over an alphabet X can be considered as an extension of the notion of a semi-Thue system over X: to each rule $u \to v$ is associated a set of words $K(u, v)$ which is the set of left-contexts where the rule can be applied. This means that a word pus can be rewritten pvs only when the left-context p belongs to $K(u, v)$. In the special case where $K(u, v)$ is the whole set X^*, we obtain the classical notion of a semi-Thue system. Let us formally define this type of rewriting system.

A *controlled rewriting system* over the alphabet X is a subset S of $X^* \times X^* \times X^*$. Every element (l, u, v) of S is called a *rule* of S. The *direct reduction generated by* S (noted $\vdash_S$) is defined as follows: for every $f, g \in X^*$, $f \vdash_S g$ iff there exist $(l, u, v) \in S$ and $s \in X^*$ such that $f = lus$ and $g = lvs$. We say that the rule (l, u, v) applies on f and leads from f to g. The relations $\overset{*}{\vdash}_S$ (reduction generated by S), $\overset{*}{\to}_S$ (derivation generated by S) are then deduced from $\vdash_S$ as described in Section 2.4.1. One can check that $\overset{*}{\leftrightarrow}_S$ is the smallest right-congruence containing the set $\{(lu, lv)\}_{(l,u,v) \in S}$. We then call $\overset{*}{\leftrightarrow}_S$ the *right-congruence generated by S.*

2.4.3. *Classes of controlled rewriting systems*

Let $\mathscr{C}_1$, $\mathscr{C}_2$ be two classes of languages. Let us denote by $\mathscr{C}_i(X)$ the set of languages over the alphabet X which belong to the class $\mathscr{C}_i$ (for $i \in \{1, 2\}$). We call $\mathscr{C}_1$-$\mathscr{C}_2$ *decomposition over* X every finite set $D = \{L_i \times \{u_i\} \times V_i\}_{i \in [1,n]}$ such that for every $i \in [1, n]$, $L_i \in \mathscr{C}_1(X)$, $u_i \in X^*$, $V_i \in \mathscr{C}_2(X)$. Each element $L_i \times \{u_i\} \times V_i$ of D is called a *component of D.* To such a decomposition D we associate the controlled rewriting system

$$\dot{D} = \bigcup_{i \in [1,n]} L_i \times \{u_i\} \times V_i.$$

We say that a controlled rewriting system S is a $\mathscr{C}_1$-$\mathscr{C}_2$ *controlled rewriting system* iff there exists a $\mathscr{C}_1$-$\mathscr{C}_2$ decomposition D such that $S = \dot{D}$. D is then said to be a $\mathscr{C}_1$-$\mathscr{C}_2$ *decomposition of S.*

We mention below some classes of controlled rewriting systems which have appeared in the literature. Let us denote by Rec, Cf, Det, Rat, Fin, respectively, the class of recursive, context-free, deterministic context-free, rational and finite language and let us use the abbreviation "c-system" for a controlled rewriting system.

The notion of a Rec-Fin c-system has been defined in [12], where the author uses this notion as a tool for testing the equivalence of simple grammars. It is noted in [14] that the systems used in [12] were Det-Fin c-systems. In [14, 15, 23], is studied the notion of a Rat-Fin c-system in relation to the notion of a context-free and deterministic context-free language. The rational semi-Thue systems, studied in [27, 28], are a special kind of Rat-Rat c-system and the context-free semi-Thue systems studied in [10] are a special kind of Rat-Alg c-system. A result of [7] shows that every Fin-Rat c-system generates a reduction which is a rational transduction.

The results obtained in the above references lead us to focus on Rat-Rat c-systems and Rat-Fin c-systems. In the following we call *rational c-system* (resp. *finite c-system*) every Rat-Rat c-system (resp. every Rat-Fin c-system). Sometimes we also use the expression "regular c-system" for rational c-system. We shall also call rational decomposition (resp. finite decomposition) every Rat-Rat decomposition (resp. every Rat-Fin decomposition).

2.4.4. Properties of controlled rewriting systems

Let S be a c-system. We shall say that S is confluent, locally confluent, Church-Rosser, noetherian, strictly l-reducing, l-reducing, strictly v-reducing, v-reducing, when the reduction $\vdash_S$ has the corresponding property (as defined in Section 2.4.1). We shall also use the notation "Irr(S)" instead of "Irr($\vdash_S$)", we shall say that a word f is S-irreducible (resp. S-reducible) when f is $\vdash_S$-irreducible (resp. $\vdash_S$-reducible).

In addition we define here some combinatorial properties which really depend on the system S (and not on the reduction $\vdash_S$ only).

2.2. Definition. Let us consider the following conditions about a c-system S:

(C1) for every rule $(r, u, v), (r', u', v') \in S$ and every $s' \in X^*$,
$ru = r'v's'$ and $|r| \leq |r'| \Rightarrow |s'| = 0$ and $|r| = |r'|$;

(C2) for every rule $(r, u, v), (r', u', v') \in S$ and every $s' \in X^*$,
$ru = r'v's'$ and $|r| > |r'| \Rightarrow |s'| = 0$ or $|r'v'| \leq |r|$;

(C3) for every rule $(r, u, v), (r', u', v') \in S$ and every $s \in X^*$,
$rus = r'v'$ and $|r| < |r'| \Rightarrow |ru| \leq |r'|$.

S is said to be left-basic iff it fulfils (C1) and (C2),
S is said to be right-basic iff it fulfils (C1) and (C3),
S is said to be basic iff it fulfils (C1), (C2) and (C3).

Each condition $C_i (i \in [1, 3])$ can be considered as the prohibition of some type of configuration involving two rules $(r, u, v), (r', u', v')$ of S.

Condition(C1): (C1) expresses that the configuration in Fig. 1(a) is impossible: $ru = r'v's'$ with $0 < |s'|$ and $|r| < |r'|$. In others words, a left-hand side of a rule cannot strictly embed a right-hand side.

Condition(C2): (C2) expresses that the configuration in Fig. 1(b) is impossible: $ru = r'v's'$ where $0 < |s'|$ and $|r'| < |r| < |r'v'|$. In other words, a left-hand side of rule cannot have a left-factor overlapped by some right-hand side of rule.

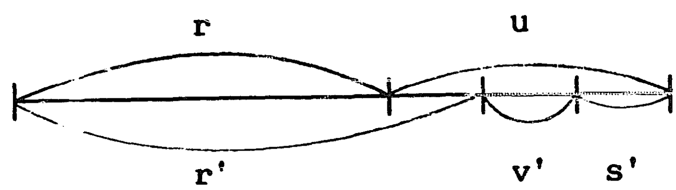

Fig. 1(a).

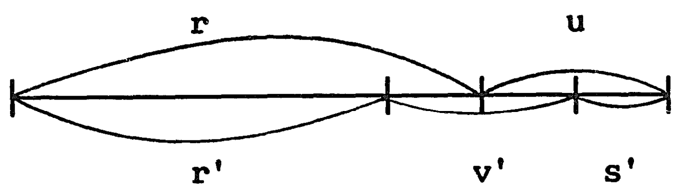

Fig. 1(b).

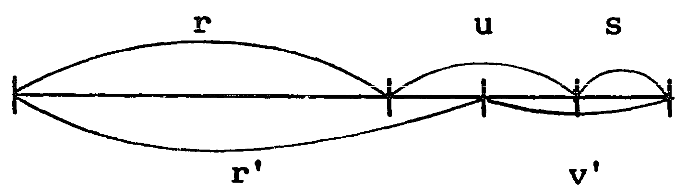

Fig. 1(c).

Condition(C3): (C3) expresses that the configuration in Fig. 1(c) is impossible: $rus = r'v'$ with $|r| < |r'| < |ru|$. In other words, a left-hand side of rule cannot have a right-factor overlapped by some right-hand side of rule.

These definitions were stated in [12, 15] and originated in earlier works about finite semi-Thue systems [25, 16, 11]. It appears that the notion of a basic system is central to the study of the links between context-free languages and congruences.

We shall also consider the following property: a c-system S is said to be *strongly injective* iff, for every rule (r, u, v), (r', u', v') of S and every $s \in X^*$, $rvs = r'v' \Rightarrow (r, u, v) = (r', u', v')$. In other words, S is strongly injective iff at most one rule can apply on a given word. This property implies that the derivation $\vdash_S$ is injective in the sense that if $f \vdash_S h$ and $g \vdash_S h$ then $f = g$. It is straightforward that every strongly-injective c-system S is confluent (because $\to_S$ is strongly confluent in the sense of [21]).

2.4.5. *Finite and rational semi-Thue systems*

A semi-Thue system over the alphabet X is, by definition, a subset T of $X^* \times X^*$. When this set T is finite, T is said to be a *finite semi-Thue system*. A semi-Thue system T over X is said to be *rational* iff it can be written as $T = \bigcup_{i=1}^{n} \{u_i\} \times V_i$ where n is some integer, each u_i is a word over X and each V_i is a rational subset of X^*. To every semi-Thue system T over X we can naturally associate a c-system $\bar{T}$ over X by setting

$$\bar{T} = \{(l, u, v) \in X^* \times X^* \times X^* \mid (u, v) \in T\}$$

or equivalently, $\bar{T} = X^* \times T$. It is then obvious that the direct reductions $\vdash_T$, $\vdash_{\bar{T}}$ are equal (where $\vdash_{\bar{T}}$ is defined in Section 2.4.2 and $\vdash_T$ is defined as usual, provided we think of $\vdash_T$ as the rewriting relation associated with a "from right to left" use of the rules). Hence, one of both systems T, $\bar{T}$ is confluent (resp. locally confluent, Church–Rosser, noetherian, strict, l-reducing, v-reducing) iff the other has the same property.

In addition we shall say that T is *left-basic* (resp. *right-basic*, *basic*) when $\bar{T}$ is left-basic (resp. right-basic, basic) in the sense of Definition 2.2. This definition of the notion of a basic semi-Thue system coincides, except for some slight details, with the definition of [32]. One can notice that every monadic or special semi-Thue system is basic (these types of semi-Thue system are investigated in several articles, see for example [18, 28, 10, 3] and see [9] for a more complete bibliography).

3. Some properties of controlled rewriting systems

3.1. Deterministic context-free languages and c-systems

The main result which motivates our study of c-systems is the following.

3.1. Theorem. *Let L be a language over a finite alphabet X. L is deterministic context-free iff there exists a regular set R over X and a c-system S over X such that S is basic, confluent, strict, finite and $L = [R]_{\overset{*}{\leftrightarrow}_S}$.*

The fact that every language $[R]_{\overset{*}{\leftrightarrow}_S}$ (where R is regular and S is a basic, confluent, strict, finite c-system) is deterministic context-free is proved in [15]. The converse is proved in [36]. This theorem remains true when we substitute the word "left-basic" for the word "basic" and (or) substitute the word "strongly injective" for the word "confluent". On one hand, the proof of [15] is given for left-basic, confluent, strict, finite c-systems. On the other hand, the proof of [36] produces a c-system which is basic, strongly injective, strict and finite.

Complement [36]. When L is prefix-free, R can be chosen finite and prefix-free.

3.2. Rational and finite c-systems

It is clear that the class of rational c-systems contains strictly the class of finite c-systems. The main result of this section is Lemma 3.6 which shows that the corresponding classes of right-congruences are equal, even when we consider the restrictions "confluent" or "strongly injective and left-basic". The consideration of rational c-systems will nevertheless be useful in Section 4.

3.2.1. A normal decomposition for strongly injective rational c-systems

3.2. Definition. A rational decomposition $D = \{(R_i, u_i, V_i)\}_{i \in [1,n]}$ is said to be normal iff

$$\forall i \in [1, n], \forall j \in [1, n], \quad R_i V_i \cap R_j V_j \neq \emptyset \Rightarrow (R_i, u_i, V_i) = (R_j, u_j, V_j)$$

3.3. Lemma. *Every strongly injective rational* c*-system has a normal rational decomposition.*

Proof. Let S be a strongly injective rational c-system defined by a rational decomposition $D_1 = \{R_i \times \{u_i\} \times V_i\}_{i \in [1,n]}$. We note $U = \{u_i\}_{i \in [1,n]}$ and for every $u \in U$ we consider

$$K(u) = \bigcup_{i \mid u_i = u} R_i \times V_i.$$

Hence $K(u)$ is a recognisable subset of $X^* \times X^*$. It is well-known that every recognisable subset of $X^* \times X^*$ can be written as a finite disjoint union of cartesian products of regular subsets of X^*. Therefore,

$$K(u) = \bigcup_{j=1}^{p(u)} H_j(u) \times W_j(u)$$

where $p(u) \in \mathbb{N}$, $H_j(u)$, $W_j(u)$ are regular subsets of X^* and if $1 \leqslant j < j' \leqslant p(u)$ then

$$(H_j(u) \times W_j(u)) \cap (H_{j'}(u) \times W_{j'}(u)) = \emptyset.$$

Let $D = \{H_j(u) \times u \times W_j(u)\}_{u \in U, 1 \leqslant j \leqslant p(u)}$ and let us check that D is normal. Let j, j' be integers and $u, u' \in U$ such that

$$H_j(u) . W_j(u) \cap H_{j'}(u') . W_{j'}(u') \neq \emptyset.$$

This means that there exist $h \in H_j(u)$, $w \in W_j(u)$, $h' \in H_{j'}(u')$, $w' \in W_{j'}(u')$ such that $hw = h'w'$. As S is strongly-injective, $(h, u, w) = (h', u', w')$. Hence $H_j(u) \times W_j(u)$ and $H_{j'}(u) \times W_{j'}(u)$ have a common element, so they must be equal. Thus,

$$(H_j(u), u, W_j(u)) = (H_{j'}(u'), u', W_{j'}(u')). \qquad \square$$

3.2.2. Transformation of a rational c*-system into an equivalent finite* c*-system*

We define here a transformation Φ, mapping every rational decomposition D into a finite decomposition $\Phi(D)$ such that $\overset{*}{\Leftrightarrow}_{\bar{D}} = \overset{*}{\Leftrightarrow}_{\Phi(D)}$. Then we establish that Φ maps every decomposition of a strict and confluent rational c-system into a decomposition of a strict and confluent finite c-system. Φ also maps every decomposition of a strongly-injective and left-basic rational c-system into a decomposition of a strongly-injective and left-basic finite c-system.

In order to define Φ we introduce new binary relations associated with a c-system. If $S \subset X^* \times X^* \times X^*$ is a c-system, we define the binary relation $\vdash_{S^r}$ over X^* by $f \vdash_{S^r} g$ iff $f = rv$ and $g = ru$ for some rule $(r, u, v) \in S$. The relations $\overset{*}{\vdash}_{S^r}$, $\overset{*}{\to}_{S^r}$ are then called, respectively, the *right-linear reduction* and the *right-linear derivation* generated by S. S is said to be *r-strongly injective* iff, for every rule (r, u, v), (r', u', v'), $rv = r'v' \Rightarrow (r, u, v) = (r', u', v')$. These definitions can be applied to a semi-Thue system T by considering the associated c-system $\bar{T} = X^* \times T$ (see Section 2.4.5).

3.4. Lemma. *For every finite* c*-system S, $\overset{*}{\vdash}_{S^r}$ is a rational transduction.*

Proof. This result is already known in the case where S is a semi-Thue system which is a recognisable subset of $X^* \times X^*$ (i.e. $S = \bigcup_{i=1}^{n} R_i \times K_i$ where each set R_i (resp. K_i) is a rational subset of X^*), see [7, 13].

Let us consider some finite c-system $S = \bigcup_{i=1}^{n} R_i \times \{u_i\} \times \{v_i\}$. We shall make the technical assumption that for every $i \in [1, n]$, $u_i \neq \varepsilon$ and $v_i \neq \varepsilon$. Let $B = \langle X, Q, q_-, d, Q \rangle$ be a finite, deterministic, complete automaton such that, for every $i \in [1, n]$, there exists a subset $Q_i^+ \subset Q$ "recognising" R_i in the sense that $R_i = L(B, Q_i^+)$. We define a new alphabet

$$Y = \{[q, x, q'] \mid q \in Q, x \in X, q' \in Q \text{ and } q.x = q'\}.$$

We define a left sequential mapping $\varphi : X^* \to Y^*$ and a homomorphism $\psi = Y^* \to X^*$ as follows:

- φ is computed by the left tranducer (see definition in [5, p. 96]), $\bar{B} = \langle X, Y, Q, q_- \rangle$, which has the same input-alphabet, set of states, initial state and transition function as B and has the following output function λ:
  $\forall q \in Q, \forall x \in X, \quad \lambda(q, x) = [q, x, q'] \quad \text{where } q' = d(q, x).$
- For every $[q, x, q']$, $\psi([q, x, q']) = x$.

We say that a word $w \in Y^*$ is monotonous iff every factor $[q_1, x_1, q_1'][q_2, x_2, q_2']$ of w is such that $q_1' = q_2$. We define a finite semi-Thue system S' by

> $S' = \{(\alpha, \beta) \in Y^* \times Y^* \mid \exists i \in [1, n]$, $\varphi(\alpha) = u_i$, $\varphi(\beta) = v_i$, α, β are monotonous and the first letter of α, β, respectively $[q_1, x_1, q_1']$, $[q_2, x_2, q_2']$, are such that $q_1 = q_2$ and $q_1 \in Q_i^+\}$.

We assert that $\vdash^*_{S^r} = \psi \circ \vdash^*_{S'^r} \circ \varphi$. hence the property that $\vdash^*_{S^r}$ is a rational transduction follows from the same property in the case of the finite semi-Thue system S'. □

We leave it to the reader to adapt this proof in the case where there exists $i \in [1, n]$ such that $u_i = \varepsilon$ or $v_i = \varepsilon$.

3.5. Lemma. *A language $L \subset X^*$ is regular iff there exists a strict, r-strongly injective finite semi-Thue system S and a finite set of words W such that $L = \langle W \rangle_{\xrightarrow{*}_{S^r}} = [W]_{\xleftrightarrow{*}_{S^r}}$.*

This lemma is proved in [17]. We restate the details of this proof because we need them in the definition of Φ.

Proof. On one hand, let us suppose that $L = \langle W \rangle_{\xrightarrow{*}_{S^r}}$ where W is finite. By Lemma 3.4, $\xrightarrow{*}_{S^r}$ (the reverse of $\vdash^*_{S^r}$) is a rational transduction. Hence L. the image of W by the rational transduction $\xrightarrow{*}_{S^r}$, is rational. On the other hand, let us suppose that L is rational. Let $\equiv_L$ be its syntactic left-congruence. Let $\preccurlyeq$ be some hierarchic order on X^*. We define a mapping $\mu_L : X^* \to X^*$ by $\mu_L(f) = g$ iff $f \equiv_L g$ and for every $h \in X^*$, $f \equiv_L h \Rightarrow g \preccurlyeq h$. Let $k_L = \max_{f \in X^*}\{|\mu_L(f)|\}$ and n_1 be an integer such that $n_1 \geqslant k_L + 1$. We define $S = \{(\mu_L(f), f) \mid f \in X^*, |f| = n_1\}$ and $W = \{w \in L, |w| \leqslant n_1 - 1\}$. It is easily verified that S is r-strongly injective and strict, that W is finite and $L = \langle W \rangle_{\xrightarrow{*}_{S^r}}$. As S is r-strongly injective, the relation $\vdash_{S^r}$ is confluent, hence $L = \langle W \rangle_{\xrightarrow{*}_{S^r}} = [W]_{\xleftrightarrow{*}_{S^r}}$. □

We are ready now to define the mapping Φ. Let $D=\{R_i\times\{u_i\}\times V_i\}_{i\in[1,n]}$ be a regular decomposition. Let $n_1=\max(\bigcup_{i=1}^{n}\{k_{v_i},|u_i|\})+1$. Let S_i and W_i be the finite semi-Thue system and the finite set of words associated with the language V_i and the integer n_1 by the proof of Lemma 3.5. Let

$$\Phi_1(D)=\{R_i\times\{u_i\}\times\{w\}\}_{i\in[1,n],w\in W_i},$$

$$\Phi_2(D)=\{R_i(V_i\alpha^{-1})\times\{\alpha\}\times\{\beta\}\}_{i\in[1,n],(\alpha,\beta)\in S_i}.$$

We define $\Phi(D)=\Phi_1(D)\cup\Phi_2(D)$.

3.6. Lemma. *Let D be a rational decomposition defining a strict rational* c*-system $\dot{D}$. Then $\Phi(D)$ is a finite decomposition defining a strict finite* c*-system $\dot{\Phi}(D)$ such that*:

(1) $\overset{*}{\leftrightarrow}_{\dot{D}}=\overset{*}{\leftrightarrow}_{\Phi(D)}$ *and* $\mathrm{Irr}(\dot{D})=\mathrm{Irr}(\dot{\Phi}(D))$;

(2) *$\dot{D}$ is confluent iff $\dot{\Phi}(D)$ is confluent*;

(3) *if $\dot{D}$ is left-basic, strongly injective and D is a normal rational decomposition, then $\dot{\Phi}(D)$ is left-basic and strongly injective.*

Proof. (1) Let us prove point (1).

(a) Let (k,g,d) be a rule of $\dot{\Phi}_1(D)$. Then (k,g,d) is also a rule of $\dot{D}$, so $kg\rightarrow_{\dot{D}}kd$. Let (k,g,d) be a rule of $\dot{\Phi}_2(D)$. Then $k\in R_i(V_i\alpha^{-1})$ and $(g,d)=(\alpha,\beta)\in S_i$ for some $i\in[1,n]$. There exist $r_i\in R_i$, $p\in V_i\alpha^{-1}$, $v_i\in V_i$, $\bar{v}_i\in V_i$ such that

$$k=r_ip,\qquad p\alpha=v_i,\qquad p\beta=\bar{v}_i.$$

Hence $kg=r_ip\alpha=r_iv_i\vdash_{\dot{D}}r_iu_i\rightarrow_{\dot{D}}r_i\bar{v}_i=r_ip\beta=kd$. Therefore $kg\overset{*}{\leftrightarrow}_{\dot{D}}kd$. We can conclude that $\overset{*}{\leftrightarrow}_{\Phi(D)}\subset\overset{*}{\leftrightarrow}_{\dot{D}}$.

(b) Let us prove the reverse inclusion. Let (r,u,v) be a rule of $\dot{D}$, which is in a component $R_i\times\{u_i\}\times V_i$ for some $i\in[1,n]$. There exists some $w_i\in W_i$ such that $v\overset{*}{\vdash}_{S_i}w_i$. Hence $rv\overset{*}{\vdash}_{\Phi_2(D)}rw_i.(r,u,w_i)$ is a rule of $\dot{\Phi}_1(D)$, so that $rw_i\vdash_{\Phi_1(D)}ru$. Finally $rv\vdash_{\Phi(D)}ru$.

We can conclude that $\vdash_{\dot{D}}\subset\overset{*}{\vdash}_{\Phi(D)}$. Together with (a) we obtain the equality $\overset{*}{\leftrightarrow}_{\dot{D}}=\overset{*}{\leftrightarrow}_{\Phi(D)}$. In addition we see that $\mathrm{Irr}(\dot{\phi}(D))\subset\mathrm{Irr}(\dot{D})$. Conversely, every word containing a redex of $\dot{\Phi}(D)$, contains a redex of $\dot{D}$, hence $\mathrm{Irr}(\dot{\Phi}(D))=\mathrm{Irr}(\dot{D})$.

(2) By Lemma 2.1, for every regular (resp. finite) c-system T which is noetherian, the property of confluence depends only on the relation $\overset{*}{\leftrightarrow}_T$ and the set $\mathrm{Irr}(T)$. $\dot{D}$ and $\dot{\Phi}(D)$ are strictly l-reducing, hence noetherian, they generate the same equivalence relation and have the same irreducible words. Hence they are either both confluent or both non-confluent.

(3) Let us suppose that $\dot{D}$ is left-basic, strongly injective and that D is a normal rational decomposition.

(a) We prove first that $\dot{\Phi}(D)$ is also strongly injective. Let (k,g,d), (k',g',d') be two rules of $\dot{\Phi}(D)$ and $s\in X^*$ such that $kds=k'd'$. We denote by $(i,j)\in\{1,2\}^2$, the case where (k,g,d) is a rule of $\dot{\Phi}_i(D)$ while (k',g',d') is a rule of $\dot{\Phi}_j(D)$.

Case 1.1: Every rule of $\dot{\Phi}_1(D)$ is a rule of $\dot{D}$. Hence $(k,g,d)=(k',g',d')$.

Case 1.2: There exist $i, j \in [1, n]$, $r_i \in R_i$, $w_i \in W_i$, $r_j \in R_j$, $v' \in X^*$, v_j, $\bar{v}_j \in V_j$ such that

$$k = r_i, \qquad g = u_i, \qquad d = w_i$$

$$k' = r_j v', \qquad v'g' = v_j, \qquad v'd' = \bar{v}_j.$$

As $\dot{D}$ is strongly injective, $(r_i, u_i, w_i) = (r_j, u_j, \bar{v}_j)$. But $w_i \in W_i \Rightarrow |w_i| \leqslant n_1 - 1$ while $|d'| = n_1$. This is in contradiction with the equality $w_i = \bar{v}_j$.

Case 2.1: By similar arguments, this case is also impossible.

Case 2.2: There exist $i, j \in [1, n]$, $r_i \in R_i$, v_i, $\bar{v}_i \in V_i$, v, $v' \in X^*$, $r_j \in R_j$, v_j, $\bar{v}_j \in V_j$ such that

$$k = r_i v, \qquad vg = v_i, \qquad vd = \bar{v}_i$$

$$k' = r_j v', \qquad v'g' = v_j, \qquad v'd' = \bar{v}_j.$$

As $\dot{D}$ is strongly injective, $(r_i, u_i, \bar{v}_i) = (r_j, u_j, \bar{v}_j)$. $|d| = |d'| = n_1$, hence $d = d'$. D is normal, so $(R_i, u_i, V_i) = (R_j, u_j, V_j)$ and hence $g = g' = \mu_{V_i}(d) = \mu_{V_j}(d')$.

(b) Let us prove now that $\dot{\Phi}(D)$ is left-basic. The conjunction of conditions (C1) and (C2) of Definition 2.2 can be summarised as: for all elementary rules (k, g, d), (k', g', d') and every word $s' \in X^*$, if $kg = k'd's'$ then $(|s'| = 0$ and $|g| \leqslant |d'|)$ or $|s'| \geqslant |g|$. Let us consider two rules of $\dot{\Phi}(D)$, (k, g, d), (k', g', d') and a word $s' \in X^*$ such that $kg = k'd's'$. As above we distinguish four cases.

Case 1.1: (k, g, d) and (k', g', d') are rules of $\dot{D}$, hence $(|s'| = 0$ and $|g| \leqslant |d|)$ or $|s'| \geqslant |g|$.

Case 1.2: There exist $i, j \in [1, n]$, $r_i \in R_i$, $w_i \in W_i$, $r_j \in R_j$, $v' \in X^*$, v_j, $\bar{v}_j \in V_j$ such that

$$k = r_i, \qquad g = u_i, \qquad d = w_i$$

$$k' = r_j v', \qquad v'g' = v_j, \qquad v'd' = \bar{v}_j.$$

As $r_i u_i = r_j v_j s'$ and S is left-basic, $(|s'| = 0$ and $|u_i| \leqslant |v_j|)$ or $|s'| \geqslant |u_i|$. As $|u_i| \leqslant n_1 - 1 < n_1 = |d'|$ we can conclude that $(|s'| = 0$ and $|u_i| < |d'|)$ or $|s'| \geqslant |u_i|$.

Case 2.1: There exist $i, j \in [1, n]$, $r_i \in R_i$, $r_j \in R_j$, $v \in X^*$, v_i, $\bar{v}_i \in V_i$, v_j, $\bar{v}_j \in V_j$ such that

$$k = r_i v \qquad vg = v_i, \qquad vd = \bar{v}_i$$

$$k' = r_j, \qquad g' = u_j, \qquad d' = w_j.$$

As S is strongly injective $(r_i, u_i, v_i) = (r_j, u_j, w_j)$ and $|s'| = 0$. As g is a suffix of d' we have $|g| \leqslant |d'|$.

Case 2.2: There exist $i, j \in [1, n]$, $r_i \in R_i$, $r_j \in R_j$, v, $v' \in X^*$, v_i, $\bar{v}_i \in V_i$, v_j, $\bar{v}_j \in V_j$ such that

$$k = r_i v, \qquad vg = v_i, \qquad vd = \bar{v}_i$$

$$k' = r_j, \qquad v'g' = v_j, \qquad v'd' = \bar{v}_j.$$

As S is strongly injective, $(r_i, u_i, v_i) = (r_j, u_j, \bar{v}_j)$ and $|s'| = 0$. As $|g| \leqslant n_i - 1$ and $|d'| = n_1$ we have $(|s'| = 0$ and $|g| \leqslant |d'|)$. $\square$

Comment. Lemma 3.6 shows that in general, strict regular c-systems have the same power as strict finite c-systems and that this result remains true when we add the restrictions "confluent" or "strongly injective and left-basic".

3.3. Leftmost reduction generated by a c-system

Let S be a c-system over the alphabet X. We call *redex* of S, every pair $(r, v) \in X^* \times X^*$ such that there exists some $u \in X^*$ such that $(r, u, v) \in S$.

A redex (r, v) is called *leftmost* iff

(1) no proper prefix of rv is S-reducible;

(2) no proper suffix v' of v is such that $rv = r'v'$ and (r', v') is a redex (we call proper prefix (resp. suffix) of a word f, every $v \in X^*$ such that there exists $u \in X^*$ with $f = vu$ (resp. $f = uv$) and such that $v \neq f$).

Let us fix some total ordering relation $\leqslant$ over X^* (for example, $\leqslant$ can be the hierarchic ordering relation deduced from some lexicographic ordering).

We say that *the rule* $(r, u, v) \in S$ *is leftmost* iff

(1) (r, v) is a leftmost redex

(2) for every rule $(r, u', v) \in S$, $u \leqslant u'$.

The *direct leftmost reduction* generated by S, noted $\vdash'_S$ is defined by: for every $f, g \in X^*$, $f \vdash'_S g$ iff there exists some leftmost rule (r, u, v) and some word $s \in X^*$ such that $f = rvs$ and $g = rus$.

We denote by $\mathrm{LMR}(S)$ the set of all leftmost rules of S. Clearly, $\mathrm{LMR}(S)$ is also a c-system and $\vdash'_S = \vdash_{\mathrm{LMR}(S)}$. In the next lemma we show that $\mathrm{LMR}(S)$ inherits some properties of S.

3.7. Lemma. (1) $\mathrm{LMR}(S)$ *is strongly injective*;

(2) *if S is left-basic (resp. basic) then* $\mathrm{LMR}(S)$ *is left-basic (resp. basic)*;

(3) *if S is strict then* $\mathrm{LMR}(S)$ *is strict*;

(4) *if S is rational then* $\mathrm{LMR}(S)$ *is rational*;

(5) *if S is finite then* $\mathrm{LMR}(S)$ *is finite*;

(6) $\mathrm{Irr}(S) = \mathrm{Irr}(\mathrm{LMR}(S))$;

(7) *if S is strict and confluent then* $\overset{*}{\leftrightarrow}_S = \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)}$.

Proof. Point (1) is straightforward. Points (2) and (3) are consequences of the inclusion $\mathrm{LMR}(S) \subset S$. In order to prove point (4) let us consider some rational decomposition $D = \{R_i \times \{u_i\} \times V_i\}_{i \in [1,n]}$ of the rational c-system S.

For every $u \in X^*$, we define several subsets of $X^* \times X^*$:

$$\mathrm{LMR}(S, u) = \{(r, v) \in X^* \times X^* \mid (r, u, v) \in \mathrm{LMR}(S)\}$$

$$E_1(S, u) = \{(r', r''v) \mid r' \in X^*, r'' \in X^+, v \in X^*, (r'r'', u, v) \in S\}$$

$$E_2(S, u) = \left\{(r, v) \mid r \in X^*, v \in X^*, rv \in \bigcup_{i=1}^{n} R_i V_i X^+\right\}$$

$$E_3(S, u) = \{(r, v) \mid \exists j \in [1, n], r \in R_j, u_j < u, v \in V_j\}.$$

One can easily check that for every $u \in X^*$,

$$\mathrm{LMK}(S, u) = \bigcup_{i \mid u_i = u} R_i \times V_i - (E_1(S, u) \cup E_2(S, u) \cup E_3(S, u)).$$

In order to prove that LMR(S) is a rational c-system, it is sufficient to prove that for every $u \in X^*$, LMK(S, u) is recognisable. By the equality above, it is sufficient to prove that for every $\alpha \in \{1, 2, 3\}$, $E_\alpha(S, u)$ is recognisable.

Case where $\alpha = 1$: For every $i \in [1, n]$, there exists an integer $l(i)$ and rational languages $H_{i,k}$ ($k \in [1, l(i)]$), $W_{i,k}$ ($k \in [1, l(i)]$), such that

$$\{(r', r'') \in X^* \times X^* \mid r'r'' \in R_i\} = \bigcup_{k=1}^{l(i)} H_{i,k} \times W_{i,k}$$

(we leave the proof of this fact to the reader). Hence

$$E_1(S, u) = \bigcup_{i \mid u_i = u} \left(\bigcup_{k=1}^{l(i)} H_{i,k} \times W_{i,k} v_i \right)$$

so that $E_1(S, u)$ is recognisable.

Cases where $\alpha = 2$ *or* $\alpha = 3$: These cases are easy and we leave them to the reader.

Hence point (4) of the lemma is proved.

In order to prove point (5), let us consider some finite c-system

$$S = \bigcup_{i=1}^{n} R_i \times \{u_i\} \times \{v_i\}.$$

By point (4), LMR(S) is rational. Hence LMR(S) $= \bigcup_{j=1}^{m} K_j \times \{u'_j\} \times W_j$ where K_j, W_j are rational. But LMR(S) $\subset S$, hence for every $j \in [1, m]$, $W_j \subset \{v_i\}_{i \in [1,n]}$ so that LMR(S) is finite.

Point (6) is true because, if some rule of S applies on a word f, then there exists some leftmost rule of S applying on f.

Let us prove point (7). It is clear that $\overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)} \subset \overset{*}{\leftrightarrow}_S$. Now, let $f, g \in X^*$ such that $f \overset{*}{\leftrightarrow}_S g$. As LMR($S$) is strict, there exist $\bar{f}, \bar{g} \in \mathrm{Irr}(\mathrm{LMR}(S))$ such that $f \vdash^*_{\mathrm{LMR}(S)} \bar{f}$ and $g \vdash^*_{\mathrm{LMR}(S)} \bar{g}$. By point (6), $\bar{f}, \bar{g} \in \mathrm{Irr}(S)$. As $\bar{f} \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)} \bar{g}$ and $\bar{f}, \bar{g} \in \mathrm{Irr}(S)$, from the confluence of S we can conclude that $\bar{f} = \bar{g}$. Hence $f \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)} g$. □

3.4. *Total reduction generated by a confluent, strict c-system*

Let S be a confluent, strict c-system over X. By Lemma 2.1, for every $f \in X^*$, $[f]_{\overset{*}{\leftrightarrow}_S} \cap \mathrm{Irr}(S)$ has one element only. Let us denote by $\theta_S(f)$ this element. We call *total reduction* of S the mapping θ_S.

Sakarovitch has shown [32] that when S is a left-basic, confluent, strict, finite semi-Thue system, θ_S preserves rationality. His proof is still available in the more general situation where S is a c-system.

3.8. Theorem. *If S is a left-basic, confluent, strict, finite controlled rewriting system over X and R is a rational subset of X^*, then $\theta_s(R)$ is also rational.*

Sketch of proof (inspired from [32]). The proof of Theorem 3 in [15], shows that there exists a dpda $A = \langle X, Y, Q, d, q_0, y_0 \rangle$ which "computes" θ_s in the sense that, if $y_0 q_0 \rightarrow^f_A wq$ then $\theta_s(f) = \varphi(w)\psi(q)$ where φ is a literal homomorphism $Y^* \rightarrow X^*$ and ψ is a mapping $Q \rightarrow X^*$. But the mapping α, sending the word f on the configuration wq reached by A when reading f, preserves rationality [20]. Hence θ_S preserves rationality. □

3.9. Remark. By Lemma 3.7, points (1), (2), (4), (7), every left-basic, confluent, strict rational c-system S defines the same total reduction as some left-basic, *strongly injective*, strict, rational c-system T (it suffices to take $T = \mathrm{LMR}(S)$). By Lemma 3.6, points (1), (3), this c-system T defines the same total reduction as some left-basic, strongly injective strict, *finite* c-system S'. Hence Theorem 3.8 remains true when we replace the notion of finite c-system by the notion of rational c-system.

4. Some decisions problems about strict finite c-systems

4.1. *Summary of the problems*

The notion of a finite c-system is a generalisation of that of a finite semi-Thue system, the notion of an equivalence relation generated by a finite c-system generalises the notion of congruence generated by a regular semi-Thue system (see Section 2.4.5 and also Lemma 3.6) and the notion of a basic (or left-basic) c-system generalises the notion of a monadic semi-Thue system. Hence, we are led to investigate, in the context of c-systems, some classical decision problems already solved (or raised) in these special cases.

Before listing these problems we mention three problems on the notions of basic and strongly injective systems.

P1: "Is S basic?"

P2: "Is S left-basic?"

P3: "Is S strongly injective?"

These three problems are decidable when S is taken in the class of strict finite c-systems. We leave the proof as an exercise (each of these problems reduces to the emptiness problem for a finite number of rational sets).

Let us list now the non-trivial problems that we investigate here.

P4: "Is S confluent?" We call P4 the *confluence problem.*

P5: "$\overset{*}{\leftrightarrow}_{S_1} \subset \overset{*}{\leftrightarrow}_{S_2}$?" When $\overset{*}{\leftrightarrow}_{S_1} \subset \overset{*}{\leftrightarrow}_{S_2}$ we say that $\overset{*}{\leftrightarrow}_{S_1}$ is a refinement of $\overset{*}{\leftrightarrow}_{S_2}$. Hence we shall call P5 the *refinement problem.*

P6: "$\overset{*}{\leftrightarrow}_{S_1} = \overset{*}{\leftrightarrow}_{S_2}$?" When $\overset{*}{\leftrightarrow}_{S_1} = \overset{*}{\leftrightarrow}_{S_2}$ we say that the systems S_1, S_2 are equivalent. Hence we shall call P6 the *equivalence problem.*

P7: "$[f]_{\overset{*}{\leftrightarrow}_{S_1}} \subset [f]_{\overset{*}{\leftrightarrow}_{S_2}}$?" We shall call P7 the *class inclusion problem.*

P8: "$[f]_{\leftrightarrow S_1} = [f]_{\leftrightarrow S_2}$?" We shall call P8 the *class equivalence problem*.

Our results are the following.

4.1. Proposition. *In the class of strict finite* c-*systems, the confluence problem is undecidable.*

This proposition is deduced from Theorem 4.1.4 of [28] which states that problem P4 is undecidable in the class of strict, regular semi-Thue systems.

4.2. Proposition. *In the class of left-basic, strict, finite* c-*systems, the confluence problem is decidable.*

In other words, P4 becomes decidable when restricted to the class of left-basic strict finite c-systems. This result generalises Theorem 3.3.6 of [28] which states that P4 is decidable when S is taken in the class of monadic strict regular semi-Thue systems.

4.3. Proposition. *In the class of pairs* (S_1, S_2) *where* S_1 *is a strict finite* c-*system and* S_2 *a left-basic, confluent, strict finite* c-*system, the refinement problem is decidable.*

This result generalises Theorem 2.5.5 of [28].

4.4. Corollary. *In the class of left-basic, confluent, strict, finite* c-*systems, the refinement problem and the equivalence problem are decidable.*

4.5. Proposition. *In the class of confluent, strict, finite* c-*systems, the equivalence problem is undecidable.*

From this result and Lemma 3.7 it follows that P6 is also undecidable in the class of strongly injective, strict, finite c-systems. For problem P7 we have no new result but we recall in the next proposition the result stated in [34, Theorem V.2].

4.6. Proposition. *In the class of confluent, basic, strict, finite semi-Thue systems, the class inclusion problem is undecidable.*

We recall that in the class of basic, confluent, strict, finite semi-Thue systems, problem P8 is decidable [34, Theorem III-2]. Using Theorem 3.1 it is not difficult to see that the general dpdas equivalence problem is equivalent to problem P8 where S_1, S_2 are taken in the class of basic (or left-basic), confluent (or strongly injective), strict, finite c-systems. (We recall that the dpdas equivalence problem consists of deciding for two dpdas A_1, A_2 and two sets of final modes F_1, F_2, whether the languages $L(A_1, F_1)$, $L(A_2, F_2)$ are equal.)

From [29, Theorem 2.1] we deduce the following.

4.7. Proposition. *In the class of confluent, strict, finite semi-Thue systems the class equivalence problem is undecidable.*

Proposition 4.7 solves a problem stated in [16, 26].

4.2. *Proofs of the results*

Proof of Proposition 4.1. Theorem 4.1.4 of [28] establishes that P4 is undecidable in the class of strict, regular semi-Thue systems. Hence, it suffices to show that this problem reduces to the same problem in the class of strict, finite c-systems.

Let $S=\bigcup_{i=1}^{n}\{u_i\}\times V_i$ be a regular, strict semi-Thue system (for every $i\in[1,n]$, V_i is regular) over a finite alphabet X. We consider $\bar{S}=\bigcup_{i=1}^{n}X^*\times\{u_i\}\times V_i$ and its decomposition,

$$\bar{D}=\{X^*\times\{u_i\}\times V_i\}_{i\in[1,n]}.$$

By Lemma 3.6, $\dot{\Phi}(\bar{D})$ is confluent iff $\dot{\bar{D}}$ is confluent. And it is obvious that $\dot{\bar{D}}$ is confluent iff S is confluent. As $\dot{\Phi}(\bar{D})$ is a strict, finite c-system, the mapping $S\to\dot{\Phi}(\bar{D})$ is the reduction we were looking for. □

We now give a series of lemmas which will allow us to prove Proposition 4.3. Proposition 4.2 will then easily follow from Proposition 4.3. We start with some definitions and notation.

4.8. Definition. Let $S=\bigcup_{i=1}^{n}R_i\times\{u_i\}\times\{v_i\}$ be a finite c-system over X. Let $m\in X^*$. The finite c-system $m^{-1}S$ is defined by

$$m^{-1}S=\bigcup_{i=1}^{n}m^{-1}R_i\times\{u_i\}\times\{v_i\}.$$

4.9. Remark. The system $m^{-1}S$ does not depend on the particular decomposition $\{R_i\times\{u_i\}\times\{v_i\}\}_{i\in[1,n]}$ and could be alternatively defined by $m^{-1}S=\{(k,u,v)\in X^*\times X^*\times X^*\mid\exists(r',u',v')\in S,\ mk=r',\ u=u',\ v=v'\}$.

4.10. Fact. *If $m=m_1m_2$ and $f\vdash^*_{m^{-1}S}g$, then $m_2f\vdash^*_{m_1^{-1}S}m_2g$.*

This fact can be easily checked. From now on up to Lemma 4.22 we fix a left-basic, strict, finite c-system $S=\bigcup_{i=1}^{n}R_i\times\{u_i\}\times\{v_i\}$ over a finite alphabet X. We define $l=\max\{|u_i|\}_{i\in[1,n]}$.

4.11. Definition. Let $(r,u)\in X^*\times X^*$. We say that (r,u) is a left-block for S iff

(i) r is S-irreducible;

(ii) for every $s'\in X^*$ and $(r',u',v')\in S$, $ru=r'v's'\Rightarrow s'=\varepsilon$.

4.12. Remark. If $u=\varepsilon$ and r is S-irreducible, (r, u) is a left-block for S. If $(r, u, v)\in S$ and r is S-irreducible, by conditions (C1), (C2) of Definition 2.2, (r, u) is a left-block for S.

4.13. Lemma. *Let r, u, s, h be words in X^* such that (r, u) is a left-block for S and $rus \stackrel{*}{\vdash}_S h$. Then there exist r_1, r_2, $w\in X^*$ such that $r=r_1r_2$, $h=r_1w$, $|w|\leqslant \max\{l, |u|\}+|s|$ and $r_2us \vdash_{r_1^{-1}S} w$.*

Proof. We prove this lemma by induction on the integer n such that $rus \vdash_S^n h$.

If $n=0$: then $r_1=r$, $r_2=\varepsilon$, $w=us$ fulfil the required properties.

Induction step: Let us suppose that $rus \vdash_S^{n+1} h$. Let us consider a word h' such that $rus \vdash_S h' \vdash_S^n h$. h' is obtained from rus by application of a rule (r', u', v'). As (r, u) is a left-block for S, one of the three following cases must occur.

Case 1 (see Fig. 2):

$$\begin{aligned} &rus = r'v's' && \text{for some } s'\in X^*,\\ &r=r'\alpha && \text{for some } \alpha\in X^*,\\ &s=\beta s' && \text{for some } \beta\in X^*. \end{aligned}$$

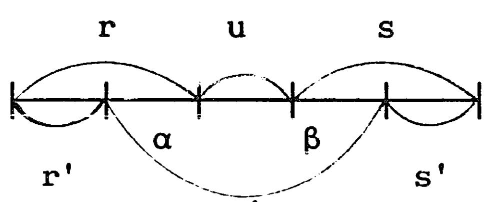

Fig. 2.

As r' is S-irreducible, by Remark 4.12 (r', u') is a left-block for S. The induction hypothesis applies on r', u', s', h. Hence there exist r_1', r_2', w' such that $r'=r_1'r_2'$, $h=r_1'w'$, $|w'|\leqslant \max\{l, |u'|\}+|s'|$ and

$$r_2'u's' \stackrel{*}{\underset{r_1'^{-1}S}{\vdash}} w'.$$

We have also

$$r_2'\alpha us = r_2'\alpha u\beta s' \underset{r_1'^{-1}S}{\vdash} r_2'u's'.$$

Hence

$$(r_2'\alpha)us \underset{r_1'^{-1}S}{\vdash} w'.$$

Moreover, $\max\{l, |u'|\}=l$ because $u'\in\{u_i\}_{i\in[1,n]}$, so that $|w'|\leqslant l\leqslant \max\{l, |u|\}$. Hence $r_1=r_1'$, $r_2=r_2'\alpha$ and $w=w'$ satisfy the required properties (see Fig. 3).

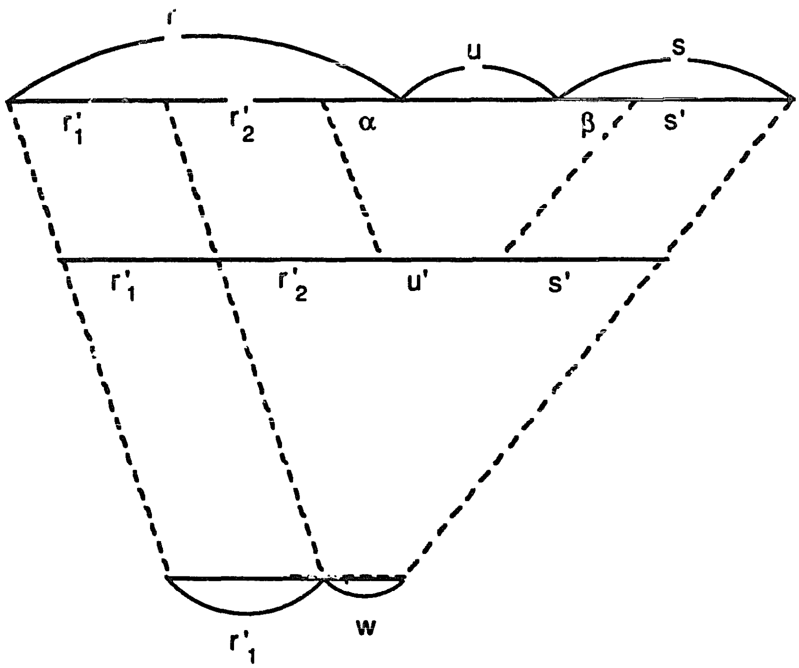

Fig. 3.

Case 2 (see Fig. 4):

$$
\begin{array}{ll}
rus = r'v's' & \text{for some } s' \in X^*,\\
r = r'\alpha & \text{for some } \alpha \in X^*,\\
v' = v_1'v_2' & \text{for some } v_1' \in X^+ \text{ and } v_2' \in X^*,\\
s = v_2's'. &
\end{array}
$$

4.14. Fact. $(r, \alpha u')$ *is a left-block for S.* Let us prove this fact. r is assumed to be S-irreducible, hence point (i) is true. Let us suppose that $r\alpha u' = r''v''s''$ for some rule $(r'', u'', v'') \in S$ and some word $s'' \in X^*$.

- If $|r\alpha| < |r''v''|$, as S is left-basic we must have $s'' = \varepsilon$ (and in addition $r\alpha = r''$).
- If $|r\alpha| \geqslant |r''v''|$, as r is S-irreducible we must have $|r| < |r''v''| \leqslant |r\alpha|$. Let s_1 be such that $r\alpha = r''v''s_1$.

Then $r\alpha v_1' = r''v''s_1v_1'$, hence $ru = r''v''(s_1v_1')$ (see Fig. 5). Either $|r| \leqslant |r''|$ and condition (C1) is violated, or $|r''| < |r| < r''v''|$ and condition (C2) is violated. Hence the fact is proved. □

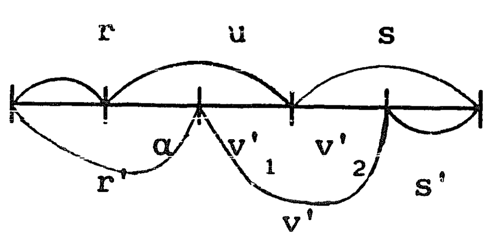

Fig. 4.

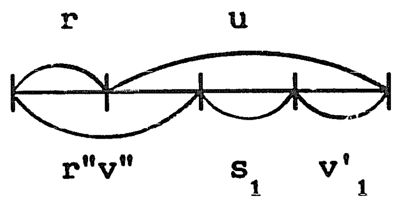

Fig. 5.

Proof of Lemma 4.13 (*continued*). Let us apply now the induction hypothesis on r, $\alpha u'$, s', h. We obtain r'_1, r'_2, w' such that $r = r'_1 r'_2$, $h = r'_1 w'$, $|w'| \leq \max\{l, |\alpha u'|\} + |s'|$ and $r'_2 \alpha u' s' \overset{*}{\vdash}_{r_1^{-1}} r'_2 w'$. Hence

$$r'_2 us \underset{r_1'^{-1}}{\overset{*}{\vdash}} r'_2 w'.$$

Moreover,

$$l + |s'| \leq l + |s| \quad \text{and} \quad |\alpha u'| + |s'| = |u| + |s|. \tag{1,2}$$

From (1) and (2) we obtain $\max\{l, |\alpha u'|\} + |s'| \leq \max\{l, |u|\} + |s|$ so that $|w'| \leq \max\{l, |u|\} + |s|$. Hence $r_1 = r'_1$, $r_2 = r'_2$ and $w = w'$ are satisfying the required properties (see Fig. 6).

Case 3 (see Fig. 7):

$$r' = ru\alpha \quad \text{for some } \alpha \in X^*$$

$$s = \alpha v' s'.$$

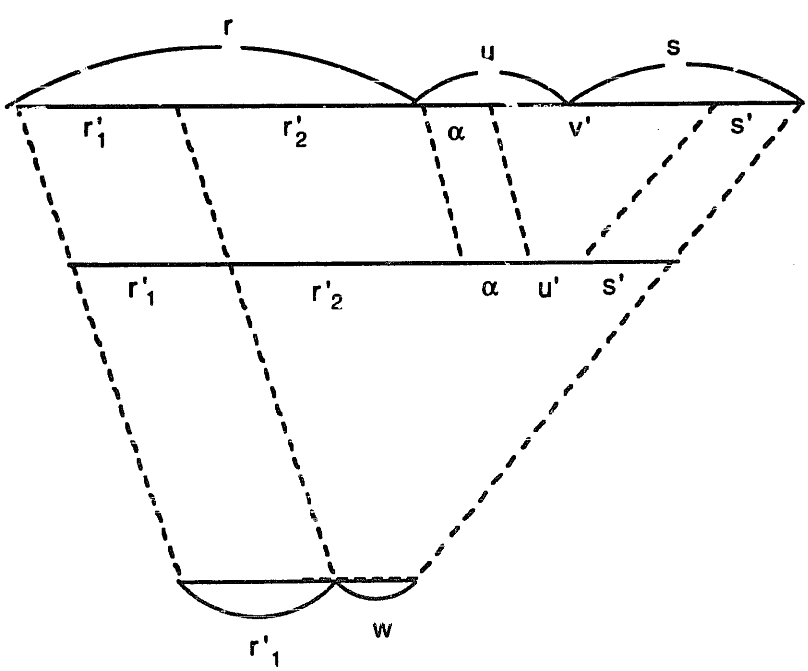

Fig. 6.

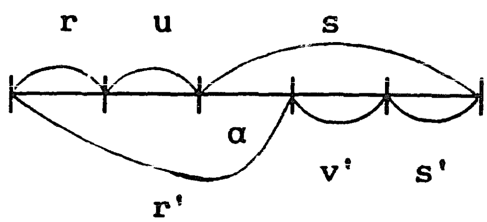

Fig. 7.

Let us apply the induction hypothesis on the words r, u, $\alpha u's'$, h. There exist r_1', r_2', w' such that

$$r = r_1'r_2', \qquad h = r_1'w', \qquad |w'| \leqslant \max\{l, |u|\} + |\alpha u's'|$$

and

$$r_2'u\alpha u's' \underset{r_1'^{-1}S}{\overset{*}{\longmapsto}} r_2'w'.$$

As $|\alpha u's'| < |s|$ it is clear that the words $r_1 = r_1'$, $r_2 = r_2'$ and $w = w'$ fulfil the required properties. □

4.15. Lemma. *Let r, f, g, h be words in X^* such that r is S-irreducible, and $rf \overset{*}{\longmapsto}_S h \overset{*}{\longleftarrow\!\!|}_S rg$. Then there exist r_1, r_2, $w \in X^*$ such that $r = r_1r_2$, $h = r_1w$, $|w| \leqslant l + \max\{|f|, |g|\}$ and $r_2f \overset{*}{\longmapsto}_{r_1^{-1}S} w \overset{*}{\longleftarrow\!\!|}_{r_1^{-1}S} r_2g$.*

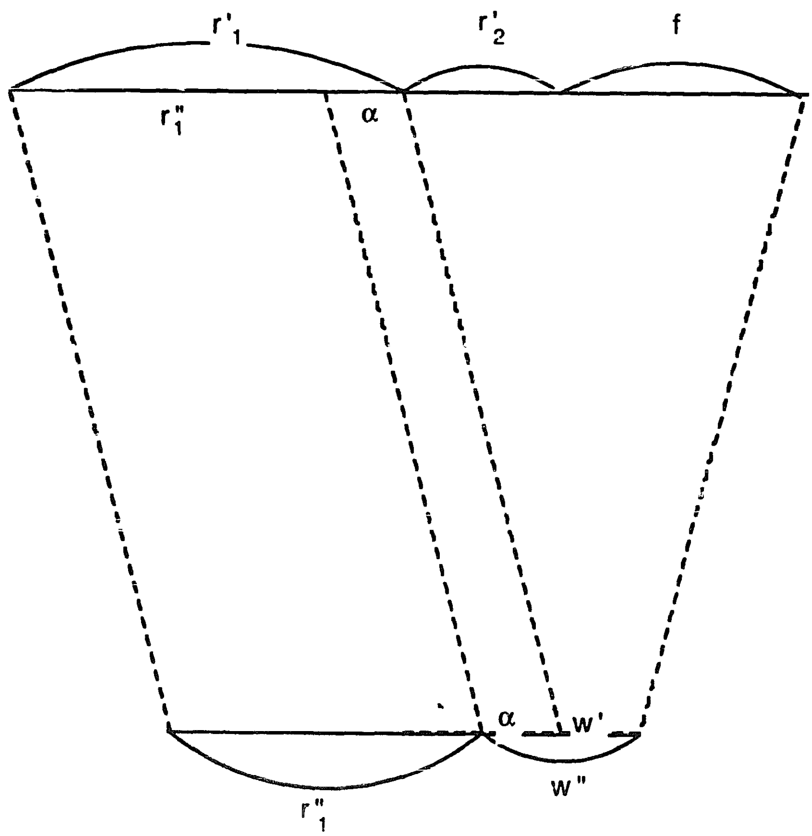

Fig. 8.

Proof. Let r, f, g, h fulfil the hypothesis. (r, ε) is a left-block for S. Hence by Lemma 4.13 there exist r_1', r_2', w' (resp. r_1'', r_2'', w'') such that $r = r_1'r_2' = r_1''r_2''$, $h = r_1'w' = r_1''w''$, $|w'| \leq l + |f|$ (resp. $|w''| \leq l + |g|$) and $r_2'f \vdash_{r_1'^{-1}S} w'$ (resp. $r_2''g \stackrel{*}{\vdash}_{r_1''^{-1}S} w''$).

Let us define the triple (r_1, r_2, w) as the element of $\{(r_1', r_2', w'), (r_1'', r_2'', w'')\}$ with the longest third component. Let us suppose, for example, that $|w'| \leq |w''|$. By our definition, $(r_1, r_2, w) = (r_1'', r_2'', w'')$. Then there exists $\alpha \in X^*$ such that $w'' = \alpha w'$, $r_1' = r_1''\alpha$ and $r_2'' = \alpha r_2'$. As $r_2'f \stackrel{*}{\vdash}_{r_1'^{-1}S} w'$, $\alpha r_2'f \stackrel{*}{\vdash}_{r_1''^{-1}S} \alpha w'$ (by Fact 4.10). Hence we have

$$r_2 f \underset{r_1^{-1}S}{\stackrel{*}{\vdash}} w \underset{r_1^{-1}S}{\stackrel{*}{\dashv}} r_2 g.$$

Moreover $|w| \leq l + |g| \leq l + \max\{|f|, |g|\}$. Hence (r_1, r_2, w) satisfies the required properties (see Fig. 8). $\square$

In the case where $|w'| > |w''|$ a symmetric argument is available.

4.16. Lemma. *Let R, K be rational subsets of X^*. Then the set $(\langle R\rangle_{\stackrel{*}{\rightarrow}_S} K^{-1}) \cap \mathrm{Irr}(S)$ is rational.*

Proof. Let us denote by H the set $(\langle R\rangle_{\stackrel{*}{\rightarrow}_S} K^{-1}) \cap \mathrm{Irr}(S)$.

$$H = \{h \in \mathrm{Irr}(S), \exists r \in R, \exists k \in K, hk \underset{S}{\stackrel{*}{\vdash}} r\}.$$

Let $A = \langle X, Q, q_-, d, Q_+\rangle$ be a deterministic finite automaton recognising the set K. We define a finite c-system T over the alphabet $Y = X \cup Q$ (we assume $X \cap Q = \emptyset$) by

$$T = \left(\bigcup_{\substack{i \in [1,n] \\ q \in Q}} R_i \times \{u_i q\} \times \{v_i q\} \right) \cup \left(\bigcup_{\substack{(q,x,q') \in Q \times X \times Q \\ q.x = q'}} \mathrm{Irr}(S) \times \{xq'\} \times \{q\} \right).$$

We assert that

$$H = \langle RQ_+\rangle_{\stackrel{*}{\rightarrow}_{T^r}} (q_-)^{-1} \tag{$\mathscr{E}$}$$

(In other words, $h \in H$ iff there exist $r \in R$ and $q \in Q_+$ such that $hq_- \stackrel{*}{\vdash}_{T^r} rq$.) This assertion can be deduced from the two following facts.

4.17. Fact. *If $r_1, r_2 \in X^*$, $q_1, q_2 \in Q$ are such that $r_1 q_1 \stackrel{*}{\vdash}_{T^r} r_2 q_2$ then there exists some $k \in X^*$ such that $q_1 . k = q_2$ and $r_1 k \stackrel{*}{\vdash}_S r_2$.*

We leave the proof of this fact to the reader. Let us mention that the fact that S is left-basic is not used in that proof.

4.18. Fact. *If $r_1, r_2, k \in X^*$, $q_1, q_2 \in Q$ are such that $q_1 . k = q_2$, $r_1 k \stackrel{*}{\vdash}_S r_2$ and r_1 is S-irreducible, then $r_1 q_1 \stackrel{*}{\vdash}_{T^r} r_2 q_2$.*

Here again, we leave the proof to the reader. Let us mention that here the hypotheses that r_1 is S-irreducible and S is left-basic are crucial. From Fact 4.17

we obtain the inclusion of the right member of ($\mathscr{E}$) in its left member. From Fact 4.18 we obtain the converse inclusion. Hence ($\mathscr{E}$) is proved. By Lemma 3.4, $\overset{*}{\vdash}_{T^r}$ (or equivalently $\overset{*}{\rightarrow}_{T^r}$) is a rational transduction. Now, R is rational and the operations product, rational transductions and right-quotient by a set preserve rationality. Hence H is rational. $\square$

4.19. Remark. In the case where S is a confluent, left-basic, strict, finite c-system, Lemma 4.16 shows that $[R]_{\overset{*}{\leftrightarrow}_S}K^{-1}=[H]_{\overset{*}{\leftrightarrow}_S}$. Hence the family of languages $\{[R]_{\overset{*}{\leftrightarrow}_S} \mid R\in \mathrm{Rat}(X^*)\}$ is a boolean algebra (this can be easily deduced from Theorem 3.8) of deterministic cfls (by Theorem 3.1) which is stable by right-quotient by rational sets.

4.20. Lemma. *Let $f, g\in X^*$ and*

$$R_S(f,g)=\{r\in \mathrm{Irr}(S), \exists h\in X^*, rf \overset{*}{\underset{S}{\vdash}} h \overset{*}{\underset{S}{\dashv}} rg\}.$$

Then $R_S(f,g)$ is a rational set.

Proof. Let us consider a right congruence $\sim$ of finite index $q\in\mathbb{N}$ which saturates every set R_i ($i\in[1,n]$). Let us denote by $\{P_j\}_{j\in[1,q]}$ the classes mod($\sim$). Let us define $p=l+\max\{|f|,|g|\}$. By Lemma 4.15, $r\in R_S(f,g)$ iff $r\in \mathrm{Irr}(S)$ and there exist w, r_1, $r_2\in X^*$ such that $r=r_1r_2$, $|w|\leqslant p$ and $r_2f \overset{*}{\vdash}_{r_1^{-1}S} w \overset{*}{\dashv}_{r_1^{-1}S} r_2g$.

It is clear that the system $r_1^{-1}S$ depends only on the class of r_1 mod($\sim$). Let us denote by $P_j^{-1}S$ the system $p_j^{-1}S$ for any element p_j of P_j. We claim that

$$R_S(f,g)=\bigcup_{\substack{|w|\leqslant p\\ j\in[1,q]}} P_j[(\langle w\rangle_{\overset{*}{\rightarrow}_{P_j^{-1}S}})f^{-1}\cap(\langle w\rangle_{\overset{*}{\rightarrow}_{P_j^{-1}S}})g^{-1}\cap \mathrm{Irr}(P_j^{-1}S)].$$

By Lemma 4.16 every term of this finite union is a rational set, hence $R_S(f,g)$ is rational. $\square$

4.21. Remark. The reader can check that the proofs of Lemmas 4.16 and 4.20 are constructive. Hence, given an effective description of S and the words f, g, one can compute a finite automaton recognising $R_S(f,g)$.

4.22. Lemma. *Let S be a left-basic, confluent, strict, finite* c*-system. Let $f, g\in X^*$ and $E_S(f,g)=\{r\in X^*, rf \overset{*}{\leftrightarrow}_S rg\}$. Then $E_S(f,g)$ is a deterministic* cfl.

Proof. As S is confluent, S is Church-Rosser, hence $E_S(f,g)=[R_S(f,g)]_{\overset{*}{\leftrightarrow}_S}$. By Lemma 4.20 $R_S(f,g)$ is rational and by Theorem 3.1, $E_S(f,g)$ is a dcfl. $\square$

4.23. Proposition. *In the class of pairs (S_1,S_2) where S_1 is a strict finite* c*-system and S_2 a left-basic, confluent, strict, finite* c*-system, the refinement problem is decidable.*

Proof. Let $S_1=\bigcup_{i=1}^n R_i\times\{u_i\}\times\{v_i\}$ and let S_2 be a left-basic, confluent, strict, finite c-system. $\overset{*}{\leftrightarrow}_{S_1}\subset\overset{*}{\leftrightarrow}_{S_2}$ iff for every $i\in[1,n]$, $R_i\subset E_{S_2}(u_i,v_i)$.

These inclusions are testable because R_i is rational while $E_{S_2}(u_i,v_i)$ is a deterministic cfl (Lemma 4.22). $\square$

4.24. Corollary. *In the class of left-basic confluent strict finite* c*-systems, the refinement problem and the equivalence problem are decidable.*

In order to show that problem P4 is decidable in the class of left-basic, strict, finite c-systems, we prove the following.

4.25. Lemma. *Let S be a left-basic, strict, finite* c*-system. S is confluent iff* $\overset{*}{\leftrightarrow}_S = \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)}$.

Proof. Let S be a left-basic, strict, finite c-system.

(1) Let us suppose that S is confluent. Then, by Lemma 3.7, point (7), $\overset{*}{\leftrightarrow}_S = \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)}$.

(2) Let us suppose that $\overset{*}{\leftrightarrow}_S = \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)}$. By Lemma 3.7, point (6), $\mathrm{Irr}(S) = \mathrm{Irr}(\mathrm{LMR}(S))$. As $\mathrm{LMR}(S)$ is confluent, using Lemma 2.1 we conclude that S is also confluent. □

4.26. Proposition. *In the class of left-basic, strict, finite* c*-systems, the confluence problem is decidable.*

Proof. Let S be a left-basic strict c-system. By Lemma 4.25 the confluence problem for S reduces to the problem: "$\overset{*}{\leftrightarrow}_S = \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)}$?". As it is always true that $\overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)} \subset \overset{*}{\leftrightarrow}_S$ we only have to decide whether $\overset{*}{\leftrightarrow}_S \subset \overset{*}{\leftrightarrow}_{\mathrm{LMR}(S)}$ or not. By Lemma 3.7 $\mathrm{LMR}(S)$ itself is a left-basic, confluent, strict, finite c-system. Hence by Proposition 4.23 the above inclusion can be tested. We have proved that problem P4 is decidable in the class of left-basic, strict, finite c-systems. □

We now want to prove the following.

4.27. Proposition. *In the class of confluent strict finite* c*-systems, the equivalence problem is undecidable.*

This result is interesting when compared with Corollary 4.24 and with the result of [24], stating that the equivalence problem is decidable in the class of confluent, strict, rational, semi-Thue systems.

The proof of this proposition, though conceptually simple, is unhappily rather long because of a number of technical details. Let us explain the main idea of this proof.

We reduce the emptiness problem for deterministic, context-sensitive languages, to the equivalence problem for two confluent, strict, rational c-systems S_1, S_2. We start with a deterministic, linear bounded Turing-machine $T = \langle X, Y, Q, d, q_0, q_f\rangle$ where X is the input-alphabet, Y is the tape-alphabet ($X \subset Y$), Q is the set of states, q_0 is the initial state, q_f is the unique final state and d, a part of $Q\times Y\times Q\times Y\times \{L, R\}$, is the set of transitions of T. We associate with T two confluent strict rational c-systems S_1, S_2 such that the following holds:

(1) The rules of S_1 simulate the transitions of T. In addition, S_1 reduces every accepting configuration of T into the symbol 1 and every halting, non-accepting configuration into the symbol 0.

(2) The rules of S_2 simulate the transitions of T, except those which apply on the initial state q_0. In contrast with S_1, S_2 reduces every configuration of T that has a state equal to q_0 into the symbol 0. As well as S_1, S_2 reduces every accepting configuration (resp. halting, non-accepting configuration) into the symbol 1 (resp. the symbol 0).

(3) In order to build strict c-systems, the simulation of the transition of T performed by S_1 and S_2 involve a special dummy symbol $\square$ (a typical simulation consists of translating the transition (q_1, x, q_2, y, R) of T by a rule $(yq_2, q_1x\square)$).

(4) S_1, S_2 are reducing most of the "incorrect words" (these are the words which do not encode any configuration of T or which contain a disposition of dummy symbols $\square$ that does not allow S_1, S_2 to simulate the next transition of T) into the symbol 0.

(5) The sets of irreducible words $\mathrm{Irr}(S_1)$, $\mathrm{Irr}(S_2)$ are equal and consist of the symbols 0, 1 and a set of "incorrect words" that the rules of S_1, S_2 cannot detect (see Lemma 4.30).

We establish that for every $f\in I-\{0,1\}$ (where $I=\mathrm{Irr}(S_1)=\mathrm{Irr}(S_2)$), $[f]_{\leftrightarrow_{S_1}} = [f]_{\leftrightarrow_{S_2}}$ (this is stated in Lemma 4.33). Hence

$$\overset{*}{\underset{S_1}{\leftrightarrow}} = \overset{*}{\underset{S_2}{\leftrightarrow}} \text{ iff } [1]_{\leftrightarrow_{S_1}} = [1]_{\leftrightarrow_{S_2}}.$$

From properties (1) and (2) of S_1, S_2 we then deduce that this last equality is true iff T recognises the empty set.

We now give a formal proof of Proposition 4.27. Let $T=\langle X, Y, Q, d, q_0, q_f\rangle$ be a deterministic linear bounded Turing-machine. From now on, the unique final state q_f is denoted by 1. We make the following technical assumptions about T.

(A1) For every q in $Q-\{1\}$ and every y in Y, there exists a unique (q', y', α) in $Q\times Y\times\{L, R\}$ such that $(q, y, q', y', \alpha)\in d$. For every y in Y, there is no (q', y', α) in $Q\times Y\times\{L, R\}$ such that $(1, y, q', y', \alpha)\in d$.

Y contains a special symbol \$, the right endmarker, which fulfils the next two assumptions.

(A2) For every q, q' in Q, y' in Y and $\alpha\in\{L, R\}$, if $(q, \$, q', y', \alpha)\in d$ then $y'=\$$ and $\alpha=L$.

(A3) There exist two states q_1, q_2 in $Q-\{q_0\}$ such that for every y in Y, $(q_0, y, q_1, \$, L)\in d$ and $(q_1, y, q_2, \$, L)\in D$.

(A4) For every (q, y, q', y', α) in d, $q'\neq q_0$.

(A5) For every (q, y, q', y', α) in d, if $q'=1$ then $\alpha=R$.

Notation: We suppose that $X \subset Y$ and $Q \cap Y = \emptyset$. Y is supposed to contain a special blank symbol B. A word uqv where $u \in Y^*$, $v \in Y^+$ and $q \in Q$ denotes the configuration of T consisting of

- the tape-content equal to uv, followed by an infinite sequence of blanks;
- the state q;
- the tape-head scanning the first letter of the word v.

We write $uqv \vdash_T u'q'v'$ iff T moves in one step from configuration uqv to configuration $u'q'v'$. The relations $\vdash_T^n$ (for some integer n), $\overset{*}{\vdash}_T$, $\overset{+}{\vdash}_T$ are then defined as usual from the initial relation $\vdash_T$.

Language recognised by T: We define the language recognised by T (noted $L(T)$) by $L(T) = \{w \in X^* \mid \exists u \in Y^*, \exists v \in Y^+, w\$q_0\$ \overset{*}{\vdash}_T u1v\}$. In other words, we say that T recognises the word w iff, starting with the word $w\$\$$ on the input-tape and with the tape-head scanning the rightmost endmarker \$, T reaches the terminal state 1.

Comments about assumptions (A1)-(A5)

(1) One can check that every deterministic context-sensitive language L such that $\varepsilon \notin L$ can be recognised by a linear bounded deterministic Turing-machine T fulfilling assumptions (A1)-(A5).

(2) By assumption (A1), T has two types of halting configurations, those of the form $u1v$ (where $u \in Y^*$, $v \in Y^+$) and those of the form qyv (where $q \in Q$, $y \in Y$, $v \in Y^*$) such that there exists some (q, y, q', y', L) in d (i.e. qyv denotes a configuration where T attempts to go to the left of the tape left-end).

(3) Assumption (A2) is just expressing that \$ is a right endmarker, that is if the tape-head scans a square on the left of a symbol \$ it will never be able either to remove this symbol \$ or to scan a square on the right of this symbol.

(4) Assumption (A3) implies that, even when starting with a configuration containing no right endmarker, T moves in two steps to a configuration where the two squares just on the right of the tape-head contain the right endmarker. This implies that if $u \in Y^*$, $y \in Y$, $v \in Y^+$ are such that $uyq_0v \overset{*}{\vdash}_T u'1v'$ for some $u' \in Y^*$, $v' \in Y^+$, then $u \in L(T)$.

(5) A consequence of assumptions (A1) and (A3) is that $q_0 \neq 1$ because some transitions start with state q_0 while there is no transition starting with state 1.

(6) Assumption (A4) implies that if $uqv \overset{+}{\vdash}_T u'q'v'$, then $q' \neq q_0$.

Let us now define the rewriting systems S_1, S_2. We define a finite alphabet A by $A = Y \cup Q' \cup \bar{Q}' \cup \{\square\} \cup \{0, 1\}$ where $Q' = Q - \{1\}$, $\bar{Q}'$ is a disjoint copy of Q' (for every $q \in Q'$, we denote by $\bar{q}$ the copy of q in $\bar{Q}'$) and $\square$ is a special new symbol.

System S_1: S_1 is the rational c-system over A defined by the decomposition D_1 made of the following components.

$$\begin{aligned} &\text{For every } (q, y, q', y', R) \in d, \quad A^* \times \{y'q'\} \times \{qy\square\} \in D_1; \\ &\text{for every } (q, y, q', y', L) \in d, \quad A^* \times \{\bar{q}'y'\} \times \{qy\square\} \in D_1 \end{aligned} \tag{1}$$

(these last components are always defined because, by (A5), $q' \neq 1$ and hence $\bar{q}'$

exists). These rules are simulating the transitions of T.

$$\text{For every } q \in Q', y \in Y, \quad A^* \times \{\square q\} \times \{q \square\square\} \in D_1 \quad \text{and}$$
$$A^* \times \{\bar{q}\square\} \times \{\square\square\bar{q}\} \in D_1 \quad \text{and} \tag{2}$$
$$A^* \times \{qy\} \times \{y\bar{q}\square\} \in D_1.$$

These rules are intended to allow the symbols $q \in Q'$ (resp. $\bar{q} \in \bar{Q}'$) to move rightwards (resp. leftwards) through the blocks of symbols $\square$.

$$\text{For every } y, y' \in Y, a \in A, q \in Q', \quad A^* \times \{0\} \times \{qyy'\} \in D_1,$$
$$A^* \times \{0\} \times \{q\square y\} \in D_1,$$
$$A^* \times \{0\} \times \{y\bar{q}y'\} \in D_1,$$
$$A^* \times \{0\} \times \{y\square\bar{q}\} \in D_1, \tag{3}$$
$$\{\varepsilon\} \times \{0\} \times \{\square\bar{q}\} \in D_1,$$
$$\{\varepsilon\} \times \{0\} \times \{\bar{q}a\} \in D_1.$$

The first four components are intended to detect the failure of the simulation of T by the system S_1. The fifth and sixth component detect that in the calculus simulated by S_1, T attempts to go on the left of the left-end of the tape.

$$\text{For every } z \in Y \cup \{\square\}, \quad A^* \times \{1\} \times \{1z\} \in D_1,$$
$$A^* \times \{1\} \times \{z1\} \in D_1;$$
$$\text{for every } t \in A, \quad A^* \times \{0\} \times \{0t\} \in D_1, \tag{4}$$
$$A^* \times \{0\} \times \{t0\} \in D_1.$$

The first two components allow S_1 to reduce every word encoding an accepting configuration of T into the symbol 1. The last two components allow S_1 to reduce every word on which S_1 either fails to simulate the calculus of T or simulates a non-accepting calculus of T, into the symbol 0.

$$A^* \times \{0\} \times A^*(\{1\} \cup Q' \cup \bar{Q}')A^*(\{1\} \cup Q' \cup \bar{Q}')A^* \in D_1. \tag{5}$$

This component reduces every word which contains at least two state-symbols (and hence does not encode any configuration of T) into the symbol 0.

System S_2: S_2 is the rational c-system over A defined by the decomposition D_2 made of the following components.

$$\text{For every } (q, y, q', y') \in Q' \times Y \times Q \times Y$$
$$\text{if } (q, y, q', y', R) \in d \text{ and } q \neq q_0, \quad A^* \times \{y'q'\} \times \{qy\square\} \in D_2, \tag{1'}$$
$$\text{if } (q, y, q', y', L) \in d \text{ and } q \neq q_0, \quad A^* \times \{\bar{q}'y'\} \times \{qy\square\} \in D_2.$$
$$A^* \times \{0\} \times A^* q_0 A^* Y A^+ \in D_2. \tag{1''}$$

The components (1′) allow S_2 to simulate the transitions of T which do not start with a state equal to q_0. The component (1″) allows S_2 to reduce almost all words encoding a configuration with state q_0 into the symbol 0. All the components of types (2), (3), (4), (5) are also in D_2. Let us show that these c-systems S_1, S_2 fulfil the required properties.

4.28. Lemma. *S_1 is strict and confluent.*

Proof. A direct inspection of all components of D_1 shows that S_1 is strict. As $\vdash_{S_1}$ is a noetherian relation, it is confluent iff it is locally confluent [21, Lemma 2-4]. Let us prove that $\vdash_{S_1}$ is locally confluent. Let $f, g, h \in A^*$ such that $f \vdash_{S_1} g$ and $f \vdash_{S_1} h$.

Case 1: $|f|_{\{1\}\cup Q'\cup \bar{Q}'} = 0$. As f is S_1-reducible, it must contain at least one occurrence of 0. Then $|g|_0 \geq 1$ and $|h|_0 \geq 1$, hence $g \overset{*}{\vdash}_{S_1} 0 \overset{*}{\dashv}_{S_1} h$.

Case 2: $|f|_{\{1\}\cup Q'\cup \bar{Q}'} \geq 2$. Then $|g|_{\{1\}\cup Q'\cup \bar{Q}'} \geq 2$ or $g \in A^*\,0\,A^*$ and $|h|_{\{1\}\cup Q'\cup \bar{Q}'} \geq 2$ or $h \in A^*\,0\,A^*$. Hence $g \overset{*}{\vdash}_{S_1} 0 \overset{*}{\dashv}_{S_1} h$.

Case 3: $|f|_{\{1\}\cup Q'\cup \bar{Q}'} = 1$. If $|f|_0 \geq 1$, as in Case 1, $g \overset{*}{\vdash}_{S_1} 0 \overset{*}{\dashv}_{S_1} h$.

Let us suppose that $|f|_0 = 0$. If $|f|_1 = 1$ then f, g, h are in $(Y \cup \{\square\})^* \{1\} (Y \cup \{\square\})^*$. Hence, using the rules of type (4), $g \overset{*}{\vdash}_{S_1} 1 \overset{*}{\dashv}_{S_1} h$. If $|f|_1 = 0$, then the redex used to reduce f in g (resp. in h) must be of type (1) or (2) or (3). One can check that two redexes of these types involving the same occurrence of a symbol in $Q' \cup \bar{Q}'$ must be equal. Hence $g = h$. $\square$

4.29. Lemma. *S_2 is strict and confluent.*

Proof. A direct inspection of the components (1′) and (1″) of D_2 shows that, as well as S_1, S_2 is strict. The considerations made in Case 1 and Case 2 of the above proof remain true for S_2. We must deal with the third case.

Case 3: $|f|_{\{1\}\cup Q'\cup \bar{Q}'} = 1$. As above, we can eliminate the case where $|f|_1 \geq 1$ and the case where $|f|_1 = 1$. Hence we suppose that $|f|_0 = |f|_1 = 0$. If the symbol of $Q' \cup \bar{Q}'$ occurring in f is different from q_0, then the argument given in the proof of Lemma 4.28 is still available for S_2. It remains to treat the case where $f = uq_0v$ for some u, $v \in (Y \cup \{\square\})^*$.

Subcase 1: $v \in \square^* Y\ (Y \cup \{\square\})^+$. Then g (resp. h) either belongs to $u\square q_0 \square^* Y\ (Y \cup \{\square\})^+$ (if the rule used to reduce f is of the type $(r, \square q_0, q_0\square\square)$) or belongs to $(Y \cup \{\square\})^*\,0\,(Y \cup \{\square\})^*$ (if the rule used to reduce f is of type (1″) or (3)). Hence $g \overset{*}{\vdash}_{S_2} 0 \overset{*}{\dashv}_{S_2} h$.

Subcase 2: $v \notin \square^* Y\ (Y \cup \{\square\})^+$. Then, the only rule that can apply on f is $(u, \square q_0, q_0\square\square)$, hence $g = h$. $\square$

4.30. Lemma. $\mathrm{Irr}(S_1) = \mathrm{Irr}(S_2) = \{0, 1\} \cup (Y \cup \{\square\})^* \cup \bar{Q}' \cup A^* Y \bar{Q}' \cup (Y \cup \{\square\})^* Q'$ $\cup (Y \cup \{\square\})^* Q' (Y \cup \{\square\})$.

Proof. (1) Let us prove first that $\mathrm{Irr}(S_1) = \mathrm{Irr}(S_2)$. It suffices to check that the sets $\mathrm{Red}(S_1)$, $\mathrm{Red}(S_2)$ are equal (where $\mathrm{Red}(S_i)$ is the set of the words which are

S_i-reducible). Let us denote by Red[j] (for $j \in [1, 5]$) the set of words that admit a redex of type (j). We define analogously Red[1'] and Red[1'']. Let $w \in$ Red[1]. Either w has a redex in $A^* \times \{qy\square\}$ where $q \in Q' - \{q_0\}$, $y \in Y$, and then $w \in$ Red[1'], or w has a redex in $A^* \times \{q_0 y\square\}$ where $y \in Y$ and then $w \in$ Red[1'']. Hence Red[1] $\subset$ Red[1'] $\cup$ Red[1''] so that

$$\mathrm{Red}(S_1) \subset \mathrm{Red}(S_2). \tag{J$_1$}$$

In order to prove the converse inclusion we have to prove that

$$\mathrm{Red}[1'] \cup \mathrm{Red}[1''] \subset \mathrm{Red}(S_1).$$

It is straightforward that Red[1'] $\subset$ Red[1]. Let $w \in$ Red[1'']. Then $w = uq_0v$ where $v \in A^*YA^+$. If v has an occurrence of a symbol in $Q' \cup \bar{Q}'$ (resp. in $\{0\}$) then v is in Red[5] (resp. Red[4]).

We suppose now that $v \in (Y \cup \{\square, 1\})^*$. If v has an occurrence of the symbol 1, as $|v| \geq 2$, v must contain a redex in $A^* \times \{z1\}$ or $A^* \times \{1z\}$ for some $z \in Y \cup \{\square\}$. Hence $v \in$ Red[4].

It remains the case where $v \in (Y \cup \{\square\})^*$. Then $v = \square^k yzv'$ for some $k \in \mathbb{N}$, $y \in Y$, $z \in Y \cup \{\square\}$, $v' \in (Y \cup \{\square\})^*$.

If $k \geq 2$ then $v \in$ Red[2],

if $k = 1$ then $v \in$ Red[3],

if ($k = 0$ and $z = \square$) then $v \in$ Red[1],

if ($k = 0$ and $z \in Y$) then $v \in$ Red[3].

We have proved that Red[1''] $\subset$ Red(S_1). Hence

$$\mathrm{Red}(S_2) \subset \mathrm{Red}(S_1) \tag{J$_2$}$$

By (J$_1$), (J$_2$) our first point is proved.

(2) Let us denote by I the set $\{0, 1\} \cup (Y \cup \{\square\})^* \cup \bar{Q}' \cup A^*Y\bar{Q}' \cup (Y \cup \{\square\})^*Q' \cup (Y \cup \{\square\})^*Q'(Y \cup \{\square\})$. One can easily check that $I \subset \mathrm{Irr}(S_1)$. It remains to prove that $A^* - I \subset \mathrm{Red}(S_1)$. Let $w \in A^* - I$. Then $w \notin (Y \cup \{\square\})^*$ so that $|w|_{\{0,1\} \cup Q' \cup \bar{Q}'} \neq 0$.

Case 1: $|w|_0 \neq 0$. $w \notin I \Rightarrow w \neq 0$, hence $w \in$ Red[4].

Case 2: $|w|_{\{1\} \cup Q' \cup \bar{Q}'} \geq 2$ and $|w|_0 = 0$. Then $w \in$ Red[5].

Case 3: $|w|_1 = 1$ and $|w|_{Q' \cup \bar{Q}'} = |w|_0 = 0$. $w \notin I \Rightarrow w \neq 1$, hence $w \in$ Red[4].

Case 4: $|w|_{Q'} = 1$ and $|w|_1 = |w|_{\bar{Q}'} = |w|_0 = 0$. Then $w = uqv$ where $q \in Q'$, u, $v \in (Y \cup \{\square\})^*$. As $w \notin I$, $|v| \geq 2$. Let us consider the prefix p of v which has length 2.

If $p \in Y\square$ then $w \in$ Red[1],

if $p = \square\square$ then $w \in$ Red[2],

otherwise $w \in$ Red[3].

Case 5: $|w|_{\bar{Q}'} = 1$ and $|w|_{Q'} = |w|_1 = |w|_0 = 0$. Then $w = u\bar{q}v$ where $q \in Q'$, u, $v \in (Y \cup \{\square\})^*$.

Subcase 1: $v=\varepsilon$. $w\notin I\Rightarrow u\in(Y\cup\{\square\})^*\square$. Hence $w=u'\square\bar{q}$ where $u'\in(Y\cup\{\square\})^*$.

If $u'=\varepsilon$ then $w\in \mathrm{Red}[3]$,

if $u'\in(Y\cup\{\square\})^*\square$ then $w\in\mathrm{Red}[2]$,

if $u'\in(Y\cup\{\square\})^*Y$ then $w\in\mathrm{Red}[3]$.

Subcase 2: $w\in\square\,(Y\cup\{\square\})^*$.

If $u=\varepsilon$ then $w\in\mathrm{Red}[3]$,

if $u=\square$ then $w\in\mathrm{Red}[3]$,

if $u\in Y$ then $w\in\mathrm{Red}[2]$.

If $|u|\geqslant 2$, either $u\in(Y\cup\{\square\})^*\square\square$ and $w\in\mathrm{Red}[2]$, or

$u\in(Y\cup\{\square\})^*Y\square$ and $w\in\mathrm{Red}[3]$, or

$u\in(Y\cup\{\square\})^+Y$ and $w\in\mathrm{Red}[2]$.

Subcase 3: $v\in Y\,(Y\cup\{\square\})^*$. If $u=\varepsilon$ or $u=\square$ or $u\in(Y\cup\{\square\})^*\square\square$ or $u\in(Y\cup\{\square\})^*Y\square$, we can conclude as in Subcase 2. If $u\in Y$ or $u\in(Y\cup\{\square\})^+Y$ then $w\in\mathrm{Red}[3]$. Hence the equality $I=\mathrm{Irr}(S_1)$ is proved. $\square$

4.31. Lemma. *For every integer $n\geqslant 0$, $m\geqslant 1$, $r\geqslant 0$, $s\geqslant 1$, every sequence of letters $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m, x'_1, x'_2, \ldots, x'_r, y'_1, y'_2, \ldots, y'_s\in Y$, every state $q, q'\in Q$ and every sequence of integers $k'_1, k'_2, \ldots, k'_r, h'_1, h'_2, \ldots, h'_s\in\mathbb{N}$, if*

$$x_1x_2\ldots x_nqy_1y_2\ldots y_m \vdash_T^p x'_1x'_2\ldots x'_rq'y'_1y'_2\ldots y'_s$$

for some $p\geqslant 0$ and $y_m=y'_s=\$$, then there exist $k_1, k_2, \ldots, k_n, h_1, h_2, \ldots, h_m\in\mathbb{N}$ such that

$$x_1\square^{k_1}x_2\square^{k_2}\ldots x_n\square^{k_n}qy_1\square^{h_1}y_2\square^{h_2}\ldots y_m\square^{h_m}$$
$$\vdash_{S_1}^* x'_1\square^{k'_1}x'_2\square^{k'_2}\ldots x'_r\square^{k'_r}q'y'_1\square^{h'_1}y'_2\square^{h'_2}\ldots y'_s\square^{h'_s}.$$

Proof. We show that this lemma is true when $p=1$. By a straightforward induction on p, one can then extend the result to every $p\geqslant 0$. We suppose that

$$x_1x_2\ldots x_nqy_1y_2\ldots y_m \vdash_T x'_1x'_2\ldots x'_rq'y'_2\ldots y'_s.$$

As $y_m=y'_s=\$$ we notice that the words

$$x_1x_2\ldots x_ny_1y_2\ldots y_m \quad\text{and}\quad x'_1x'_2\ldots x'_ry'_1y'_2\ldots y'_s$$

are equal.

Case 1: The transition used is $qy_1\vdash_T x'_rq'$. We notice that $qy_1\square^{2k'_r+1}y'_1\vdash_{S_1}^* x'_r\square^{k'_r}q'y'_1$. Hence we set $k_1=k'_1,\ldots,k_n=k'_{r-1}$, $h_1=2k'_r+1$, $h_2=h'_1,\ldots,h_m=h'_s$. These integers $k_1,\ldots,k_n$, $h_1,\ldots,h_m$ satisfy the lemma.

Case 2: The transition used is $x_n q y_1 \vdash_T q' y_1' y_2'$ (where $x_n = y_1'$). We notice that $x_n \,\square^{(2h_1'+2)}\, q y_1 \,\square^{(h_2'+1)} \overset{*}{\vdash}_{S_1} q' y_1' \,\square^{h_1'}\, y_2' \,\square^{h_2'}$. Hence we set $k_1 = k_1', \ldots, k_n = 2h_1'+2$, $h_1 = h_2'+1$, $h_2 = h_3', \ldots, h_m = h_s'$. These integers $k_1, \ldots, k_n$, $h_1, \ldots, h_m$ satisfy the lemma. □

4.32. Lemma. *Lemma* 4.31, *where Q is replaced by $Q-\{q_0\}$ and S_1 by S_2, remains true.*

Proof. By assumption (A4), the state q_0 cannot arise in a computation of T starting with a state $q \neq q_0$. As every transition of T, which does not involve state q_0, can be simulated by S_2 as well as S_1, the assertion of Lemma 4.31 remains true for a computation starting with a state $q \neq q_0$ and for the system S_2. □

4.33. Lemma. *For every f in $I-\{0,1\}$, $[f]_{\leftrightarrow_{S_1}} = [f]_{\leftrightarrow_{S_2}}$ (where I denotes the set* $\mathrm{Irr}(S_1) = \mathrm{Irr}(S_2)$).

Proof. We prove that for every $w \in A^*$,

(a) $\theta_{S_1}(w) \in \{0,1\}$ iff $\theta_{S_2}(w) \in \{0,1\}$;

(b) if $\theta_{S_1}(w) \notin \{0,1\}$ then $\theta_{S_1}(w) = \theta_{S_2}(w)$.

Let us consider several cases.

Case 1: $|w|_{\{q_0,\bar{q}_0\}} = 0$. Let us show by induction on $|w|$ that in this case $\theta_{S_1}(w) = \theta_{S_2}(w)$.

If $|w| = 0$, then $w \in I$ and $\theta_{S_1}(w) = \theta_{S_2}(w) = \varepsilon$,

if $|w| = n+1$ and $w \in I$, then $\theta_{S_1}(w) = \theta_{S_2}(w) = w$,

if $|w| = n+1$ and $w \notin I$, then there exists w' such that $w \underset{S_1}{\vdash} w'$.

As the rule leading from w to w' does not involve the symbol q_0, we have also $w \vdash_{S_2} w'$. By assumption (A4), $|w'|_{\{q_0,\bar{q}_0\}} = 0$ and as S_1 is strict $|w'| \leqslant n$. By the induction hypothesis, $\theta_{S_1}(w') = \theta_{S_2}(w')$. For every $i \in \{1,2\}$, $\theta_{S_i}(w) = \theta_{S_i}(w')$, hence $\theta_{S_1}(w) = \theta_{S_2}(w)$.

Case 2: $|w|_{q_0} = 1$. If $|w|_{\{1\}\cup Q'\cup \bar{Q}'} \geqslant 2$ or $|w|_0 \geqslant 1$, then $\theta_{S_1}(w) = \theta_{S_2}(w) = 0$. Hence we can suppose that $w = u q_0 v$ for some u, $v \in (Y \cup \{\square\})^*$.

Subcase 1: $v \in \square^*$. $v = \square^{2k+r}$ for some $k \geqslant 0$ and r such that $0 \leqslant r \leqslant 1$. Then $\theta_{S_1}(w) = \theta_{S_2}(w) = u \,\square^k q_0 \,\square^r$.

Subcase 2: $v \in \square^* Y$. $v = \square^{2k+r} y$ where $k \geqslant 0$, $0 \leqslant r \leqslant 1$ and $y \in Y$.

If $r = 0$ then $\theta_{S_1}(w) = \theta_{S_2}(w) = u \,\square^k q_0 y$,

if $r = 1$ then $\theta_{S_1}(w) = \theta_{S_2}(w) = 0$.

Subcase 3: $v \in \square^* Y(Y \cup \{\square\})^+$. Then $w \vdash_{S_2} 0$ (using a rule of type (1″)). Hence $\theta_{S_2}(w) = 0$. We have to prove that $\theta_{S_1}(w) \in \{0,1\}$. $v = \square^{2k+r} y y' v'$ where $k \geqslant 0$, $0 \leqslant r \leqslant 1$, $y \in Y$, $y' \in Y \cup \{\square\}$, $v' \in (Y \cup \{\square\})^*$.

If $r = 1$ then $\theta_{S_1}(w) = 0$,

if $r = 0$ then $w \underset{S_1}{\overset{*}{\vdash}} u \,\square^k q_0 y y' v'$.

Let us use assumption (A3). If $y' \in Y$ then

$$u \,\square^k q_0 y y' v' \underset{S_1}{\vdash} u \,\square^k 0 v' \underset{S_1}{\overset{*}{\vdash}} 0,$$

hence $\theta_{S_1}(w) = 0$; else

$$u \,\square^k q_0 y y' v' \underset{S_1}{\vdash} u \,\square^k \bar{q}_1 \$ v'.$$

If $u \in \square^*$ then

$$u \,\square^k \bar{q}_1 \$ v' \underset{S_1}{\overset{*}{\vdash}} 0,$$

hence $\theta_{S_1}(w) = 0$; else $u = u' y'' \square^l$ for some $u' \in (Y \cup \{\square\})^*$, $y'' \in Y$, $l \geq 0$. If $k + l$ is odd or $k + l = 0$ then

$$u' y'' \square^{k+l} \bar{q}_1 \$ v' \underset{S_1}{\overset{*}{\vdash}} 0,$$

hence $\theta_{S_1}(w) = 0$; else $k + l = 2k' \geq 2$ and

$$u' y'' \square^{k+l} \bar{q}_1 \$ v' \underset{S_1}{\overset{*}{\vdash}} u' q_1 y'' \square^{k'-1} \$ v'.$$

If $k' = 1$,

$$u' q_1 y'' \$ v' \underset{S_1}{\vdash} u' 0 v' \underset{S_1}{\overset{*}{\vdash}} 0,$$

hence $\theta_{S_1}(w) = 0$; else

$$u' q_1 y'' \square^{k'-1} \$ v' \underset{S_1}{\vdash} u' \bar{q}_2 \$ \square^{k'-2} \$ v' = w'.$$

Because of assumption (A2), it is not possible that

$$\theta_{S_1}(w') \in \bar{Q}' \cup A^* Y \bar{Q}' \cup (Y \cup \{\square\})^* Q' \cup (Y \cup \{\square\})^* Q' (Y \cup \{\square\}).$$

Hence $\theta_{S_1}(w') \in \{0, 1\}$ so that $\theta_{S_1}(w) \in \{0, 1\}$.

Case 3: $|w|_{\bar{q}_0} = 1$. As in Case 2, we can suppose that $w = u \bar{q}_0 v$ for some u, $v \in (Y \cup \{\square\})^*$. Then, either

$$\theta_{S_1}(w) = \theta_{S_2}(w) \in \{\bar{q}_0\} \cup A^* Y \bar{q}_0 \cup \{0\} \quad \text{or} \quad w \underset{S_1}{\overset{*}{\vdash}} u' q_0 v' = w',$$

where this reduction uses only rules of type (2). Hence we have also $w \overset{*}{\vdash}_{S_2} w'$. Now, by the result of Case 2, w' fulfils (a) and (b). Hence w fulfils (a) and (b).

Case 4: $|w|_{\{q_0, \bar{q}_0\}} \geq 2$. Then, the rules of type (5) ensure that $\theta_{S_1}(w) = \theta_{S_2}(w) = 0$.

4.34. Lemma. $\overset{*}{\leftrightarrow}_{S_1} = \overset{*}{\leftrightarrow}_{S_2}$ *iff* $[1]_{\leftrightarrow_{S_1}} = [1]_{\leftrightarrow_{S_2}}$.

Proof. $\overset{*}{\leftrightarrow}_{S_1} = \overset{*}{\leftrightarrow}_{S_2}$ iff, for every $f \in I$, $[f]_{\leftrightarrow_{S_1}} = [f]_{\leftrightarrow_{S_2}}$. By Lemma 4.33 this is true if $[1]_{\leftrightarrow_{S_1}} = [1]_{\leftrightarrow_{S_2}}$. $\square$

4.35. Lemma. *If* $L(T) \neq 0$ *then* $[1]_{\leftrightarrow_{S_1}} - [1]_{\leftrightarrow_{S_2}} \neq \emptyset$.

Proof. Let $w \in L(T)$, $w = x_1 x_2 \ldots x_n$ (where each x_i is a letter in X). By definition of $L(T)$, $x_1 x_2 \ldots x_n \$ q_0 \$ \overset{*}{\vdash}_T u 1 v$ for some u, $v \in Y^*$. By assumption (A2), $v = v' \$$

for some $v' \in Y^*$ hence $x_1x_2 \ldots x_n\$q_0\$ \overset{*}{\vdash}_T u1v'\$$. By Lemma 4.31, there exist h_1, $h_2, \ldots, h_n, h_{n+1}, h_{n+2}$ such that

$$x_1 \square^{h_1} x_2 \square^{h_2} \ldots x_n \square^{h_n} \$ \square^{h_{n+1}} q_0\$ \square^{h_{n+2}} \underset{S_1}{\overset{*}{\vdash}} u1v'\$.$$

Moreover, $u1v'\$ \overset{*}{\vdash}_{S_1} 1$ (using rules of type (4)). Hence the word

$$w' = x_1 \square^{h_1} X_2 \square^{h_2} \ldots x_n \square^{h_n} \$ \square^{h_{n+1}} q_0\$ \square^{h_{n+2}}$$

is in $[1]_{\leftrightarrow_{S_1}}$. Let us notice that $h_{n+2} \geqslant 1$, otherwise w' would be S_1-irreducible. Hence $w' \in A^*q_0A^*YA^+$, and by means of a rule of type (1″) we have $w' \vdash_{S_2} 0$. Hence $\theta_{S_2}(w') \neq 1$ and $w' \in [1]_{\leftrightarrow_{S_1}} - [1]_{\leftrightarrow_{S_2}}$. □

4.36. Lemma. *If* $L(T) = \emptyset$ *then* $[1]_{\leftrightarrow_{S_1}} = [1]_{\leftrightarrow_{S_2}}$.

Proof. Let us suppose that $L(T) = \emptyset$. We prove that, under this hypothesis, for every $w \in A^*$,

$$[\theta_{S_1}(w) = 1 \Leftrightarrow \theta_{S_2}(w) = 1].$$

Let us follow the same schema as in the proof of Lemma 4.33.

Case 1: $|w|_{\{q_0, \bar{q}_0\}} = 0$. Then, as shown in the proof of Lemma 4.33, $\theta_{S_1}(w) = \theta_{S_2}(w)$.

Case 2: $|w|_{q_0} = 1$. We can suppose that $w = uq_0v$ for some $u, v \in (Y \cup \{\square\})^*$. By the conclusions obtained in Subcase 1, Subcase 2 of the proof of Lemma 4.33, we only have to consider Subcase 3, that is

$$v = \square^{2k+r}yy'v' \quad \text{where } k \geqslant 0,\ 0 \leqslant r \leqslant 1,\ y \in Y,\ y' \in Y,\ v' \in (Y \cup \{\square\})^*.$$

The considerations made in Subcase 3 show that $\theta_{S_2}(w) = 0$, $\theta_{S_1}(w) \in \{0, 1\}$ and the only case where it is possible that $\theta_{S_1}(w) = 1$ is $r = 0$, $y' = \square$, $u = u'y''\square^l$, $k + l = 2k' \geqslant 4$ where $u' \in (Y \cup \{\square\})^*$, $y'' \in Y$. Then

$$w = u'y'' \square^l q_0 \square^{2k} y \square v' \underset{S_1}{\overset{*}{\vdash}} u'\bar{q}_2\$ \square^{k'-2} \$v' = w'.$$

Let us define $w'' = u'\$ \square^l q_0 \square^{2k} \$ \square v'$. $w'' \overset{*}{\vdash}_{S_1} w' \overset{*}{\dashv}_{S_1} w$ so that $\theta_{S_1}(w) = \theta_{S_1}(w'')$. Let us denote by u'' the word u' where we delete all occurrences of the symbol $\square$. If $\theta_{S_1}(w) = 1$, then $\theta_{S_1}(w'') = 1$ which shows that $u'' \in L(T)$, contradicting the hypothesis that $L(T) = \emptyset$. Hence $\theta_{S_1}(w) = \theta_{S_2}(w) = 0$.

Case 3: $|w|_{\bar{q}_0} = 1$. As in the proof of Lemma 4.33, this case reduces to case 2.

Case 4: $|w|_{\{q_0, \bar{q}_0\}} \geqslant 2$. It is clear that $\theta_{S_1}(w) = \theta_{S_2}(w) = 0$.

Hence, in all cases, the proposition $[\theta_{S_1}(w) = 1 \Leftrightarrow \theta_{S_2}(w) = 1]$ is true. □

Proof of Proposition 4.5. By Lemmas 4.35 and 4.36, the emptiness problem for the deterministic context-sensitive language L reduces to the problem "$[1]_{\leftrightarrow_{S_1}} = [1]_{\leftrightarrow_{S_2}}$?". By Lemma 4.33, this problem is equivalent to the problem "$\overset{*}{\leftrightarrow}_{S_1} = \overset{*}{\leftrightarrow}_{S_2}$?". By Lemmas 4.28 and 4.29, S_1, S_2 are confluent, strict, c-systems. As the emptiness problem for deterministic context-sensitive languages is undecidable, we can conclude that the equivalence problem is undecidable in the class of confluent, strict, rational c-systems. Finally, using Lemma 3.6, we can replace the notion of a rational c-system by the notion of a finite c-system in the above statement. □

4.37. Remark. As mentioned at the beginning of Section 4, problem P6 is also undecidable in the class of strongly injective strict finite c-systems (this follows from Lemma 3.7).

The details of the proof of Proposition 4.5 show that given two confluent strict finite c-systems S_1, S_2 and two words $f, g \in \mathrm{Irr}(S_1)$, even when we have the information that $\mathrm{Irr}(S_1) = \mathrm{Irr}(S_2)$, that for every $h \in \mathrm{Irr}(S_1) - \{f, g\}$, $[h]_{\leftrightarrow_{S_1}} = [h]_{\leftrightarrow_{S_2}}$ and that $[f]_{\leftrightarrow_{S_2}} \subset [f]_{\leftrightarrow_{S_1}}$, we cannot decide whether $[f]_{\leftrightarrow_{S_2}} = [f]_{\leftrightarrow_{S_1}}$ or not.

The above remark shows that P8 is undecidable in the class of confluent strict finite c-systems. Proposition 4.38 is an extension of this result to the class of confluent strict finite semi-Thue systems.

4.38. Proposition. *In the class of confluent, strict, finite semi-Thue systems, the class equivalence problem is undecidable.*

Proof. To prove Proposition 4.38, we use the following theorem proved in [29], using a construction of [27].

4.39. Theorem (Otto [29]). *Let $L \subset \Sigma^*$ be recursively enumerable, and let \$, s_0, ¢ be new letters. Then there are a finite alphabet $\Gamma \supset \Sigma \cup \{\$, s_0, ¢\}$, a confluent strict finite semi-Thue system T over Γ and a regular subset R of Γ^* such that*

$$\theta_T(R) \cap \{\$s_0\}\Sigma^*\{¢\} = \Delta_T^*(R) \cap \{\$s_0\}\Sigma^*\{¢\} = [R]_{\leftrightarrow_T} \cap \{\$s_0\}\Sigma^*\{¢\} = \{\$s_0\}L\{¢\}.$$

Here, for every subset A of Γ^*, $\Delta_T^*(A)$ is defined by

$$\Delta_T^*(A) = \{w \in \Gamma^* \mid \exists a \in A, a \overset{*}{\underset{T}{\vdash}} w\}.$$

The first of the three equalities stated in this version of Theorem 4.39 is not explicit in the formulation of [29] but is a straightforward consequence of a property of T established in the proof of [29, p. 268]; property (d) states that $\{\$s_0\}.\Sigma^*.\{¢\}$ is included in $\mathrm{Irr}(T)$.

From this theorem we deduce that one can compute an alphabet X, a confluent strict semi-Thue system T over X and a regular set R such that $\theta_T(R)$ is recursively enumerable but not recursive (it suffices to consider a language L which is recursively enumerable but not recursive and to apply Theorem 4.39). In addition, we can suppose that $\varepsilon \notin R$, because if $\theta_T(R)$ is non-recursive, $\theta_T(R - \{\varepsilon\})$ is also non-recursive.

From now on, up to the end of Section 4, the alphabet X, the system S and the rational set R are fixed. Let $B = \langle X, Q, q_-, d, Q_+\rangle$ be a determinstic complete finite automaton recognising the rational set R. We define a new alphabet Y composed of the following symbols:

- all the letters $x \in X$

- a disjoint copy $\hat{X}$ of the alphabet X (each letter $x \in X$, has a copy in $\hat{X}$ denoted by $\hat{x}$)
- three special symbols $\#$, $\underline{\#}$, $\bar{\#}$
- all the symbols $q \in Q$.

We assume, of course, that the alphabets X, $\hat{X}$, Q and $\{\#, \underline{\#}, \bar{\#}\}$ are pairwise disjoint. We define two finite semi-Thue systems S_1, S_2 over Y.

S_1 is composed of the following rules:

(1) $(\#xq, \#\hat{x})$ for every $x \in X$, $q \in Q$ such that $q.x = q$,

(2) $(xq', q\hat{x})$ for every $x \in X$, q, $q' \in Q$ such that $q.x = q'$,

(3) $(\bar{\#}, q_+\underline{\#})$ for every $q_+ \in Q_+$,

(4) (u, v) for every $(u, v) \in T$.

S_2 is composed of all the rules of type (2), (3), (4). In other words, S_2 is equal to the system S_1 where we have removed the rules of type (1).

Given a word $f \in X^*$, by $\hat{f}$ we denote its copy on $\hat{X}$. Given a language $L \subset X^*$, by $\hat{L}$ we denote the set $\{\hat{f} \mid f \in L\}$.

In the next two lemmas we describe the words that belong to the sets $\langle \# w \bar{\#} \rangle_{\overset{*}{\leftarrow}_{S_1}}$ or $\langle \# w \bar{\#} \rangle_{\overset{*}{\leftarrow}_{S_2}}$ for some $w \in X^*$. Lemma 4.42 then shows that the membership problem for a word $w \in \mathrm{Irr}(T)$ in $\theta_T(R)$ reduces to the class equivalence problem for S_1, S_2 on the word $\# w \bar{\#}$.

4.40. Lemma. *Let $w \in X^*$. A word $f \in Y^*$ is such that $f \overset{*}{\vdash}_{S_1} \# w \bar{\#}$ iff f has one of the following decompositions*:

(1) $f = \# h \bar{\#}$ *where* $h \in X^*$ *and* $h \overset{*}{\vdash}_T w$;

(2) $f = \# h_1 q \hat{h}_2 \underline{\#}$ *where* h_1, $h_2 \in X^*$, $q.h_2 \in Q_+$ *and* $h_1 h_2 \overset{*}{\vdash}_T w$;

(3) $f = \# \hat{h} \underline{\#}$ *where* $h \in R$ *and* $h \overset{*}{\vdash}_T w$.

4.41. Lemma. *Let $w \in X^*$. A word $f \in Y^*$ is such that $f \overset{*}{\vdash}_{S_2} \# w \bar{\#}$ iff f has one of the following decompositions*:

(1) $f = \# h \bar{\#}$ *where* $h \in X^*$ *and* $h \overset{*}{\vdash}_T w$;

(2) $f = \# h_1 q \hat{h}_2 \underline{\#}$ *where* h_1, $h_2 \in X^*$, $q.h_2 \in Q_+$ *and* $h_1 h_2 \overset{*}{\vdash}_T w$.

Lemmas 4.40 and 4.41 can be proved by induction on the integer n such that $f \overset{n}{\vdash}_{S_i} \# w \bar{\#}$ $(i \in \{1, 2\})$. We omit this proof which is straightforward.

4.42. Lemma. *Let $w \in X^*$ such that w is T-irreducible. Then $w \in \theta_T(R)$ iff $[\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_1}} \neq [\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_2}}$.*

Proof. Let $w \in \mathrm{Irr}(T)$. As $S_2 \subset S_1$, $[\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_2}} \subset [\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_1}}$. By Lemmas 4.40 and 4.41, $[\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_2}} \neq [\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_1}}$ iff $[\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_1}}$ contains at least one word f of type (3). In that case $f = \# \hat{h} \underline{\#}$ where $h \in R$ and $h \overset{*}{\vdash}_T w$, which shows that $w \in \theta_T(R)$. Conversely, if $w \in \theta_T(R)$, then there exists $h \in R$ such that $h \overset{*}{\vdash}_T w$, hence $[\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_1}}$ contains the word $f = \# \hat{h} \underline{\#}$ which is of type (3). We conclude that $[\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_2}} \neq [\# w \bar{\#}]_{\overset{*}{\leftrightarrow}_{S_1}}$ iff $w \in \theta_T(R)$. $\square$

4.43. Proposition. *There exist two confluent strict finite semi-Thue systems T_1, T_2 over a finite alphabet Y such that $T_1 \subset T_2$ and the function $E: Y^* \to \{0, 1\}$ defined by*

$$E(f) = \textbf{if } [f]_{\leftrightarrow_{T_1}} = [f]_{\leftrightarrow_{T_2}} \textbf{ then } 1 \textbf{ else } 0$$

is not recursive.

Proof. Let us consider the semi-Thue systems S_1, S_2 defined above. By Lemma 4.42 the function E associated with S_1, S_2 is not recursive because the set $\theta_T(R)$ is not recursive. Let us consider the valuation $\nu: Y^* \to \mathbb{N}$ defined by

$$\forall y \in X \cup Q \cup \{\#, \bar{\#}, \not\#\}, \qquad \nu(y) = 1,$$

$$\forall y \in \hat{X} \qquad \nu(y) = 3.$$

S_1 and S_2 are v-strict with respect to this valuation ν. S_1 (resp. S_2) has the same set of critical pairs as T. As T is locally confluent, S_1 (resp. S_2) is locally confluent. But S_1 (resp. S_2) is v-strict, hence it is confluent. In order to have strict semi-Thue systems we consider three copies Y_1, Y_2, Y_3 of Y and an homomorphism $\varphi: Y^* \to (Y_1 \cup Y_2 \cup Y_3)^*$ such that

$$\forall y \in X \cup Q \cup \{\#, \bar{\#}, \not\#\}, \quad \varphi(y) = y_1 \quad \text{(its copy on } Y_1\text{)},$$

$$\forall y \in \hat{X} \qquad \varphi(y) = y_1 y_2 y_3$$

and we define

$$T_1 = \{(\varphi(u), \varphi(v)) \mid (u, v) \in S_1\},$$

$$T_2 = \{(\varphi(u), \varphi(v)) \mid (u, v) \in S_2\}$$

It is clear that T_1, T_2 are confluent and have an associated function E which is non-recursive. In addition T_1, T_2 are strict. Hence T_1, T_2 are two confluent, strict, finite semi-Thue systems such that E is not recursive. □

Proof of Proposition 4.38 (*conclusion*). From Proposition 4.43, Proposition 4.38 follows immediately. □

5. Decision problems about finite semi-Thue systems in connection with the equivalence problem for dpda

In this section we use the notion of a c-system as a tool allowing us to obtain results about the classical notion of a semi-Thue systems. In Section 5.1 we give a representation theorem which characterises deterministic context-free languages in terms of left-basic, confluent, finite semi-Thue systems. In Section 5.2 we show that the classical equivalence problem for dpda is interreducible with three decision problems concerning left-basic, strict, finite semi-Thue systems (Theorem 5.17). Among these three problems is the problem of deciding whether a semi-Thue system S is partially confluent on a given word f (i.e. $[f]_{\leftrightarrow_S} = \langle f \rangle_{\leftrightarrow_S}$).

We then show that this problem is decidable in the class of basic, strict, finite semi-Thue systems (Theorem 5.26) and that it is undecidable in the class of strict, finite semi-Thue systems (proposition 5.31).

5.1. A representation of dcfls by left-basic, confluent, finite semi-Thue systems

Our first aim is to prove the following.

5.1. Proposition. *Let L be a language over the finite alphabet X and let $\#$, $\$$ be new letters. L is deterministic context-free, iff there exists an alphabet Y containing $X \cup \{\#, \$\}$, a word $f \in Y^*$ and a semi-Thue system S over Y such that S is left-basic, confluent,* v*-strict, finite and*

$$\# L\$ = [f]_{\leftrightarrow_S} \cap (X \cup \{\#, \$\})^*.$$

This statement is similar to Theorem 3.1. It is a kind of converse to Theorem 7-6 of [32] (which generalises the theorem of [25]) which states that if S is left-basic, confluent, strict, finite then $[f]_{\leftrightarrow_S}$ is a dcfl. This Proposition 5.1 is also analogous to a theorem of [29] which gives a representation of recursively enumerable languages by confluent, strict, finite semi-Thue systems (it is cited here as Theorem 4.39). As remarked above, the "if" part of Proposition 5.1 is a consequence of Theorem 7-6 of [32]. It remains to prove that given a dcfl L, one can find Y, S and f as asserted in the proposition. Let us fix some dcfl L over a finite alphabet X. $\# L\$$ is a prefix-free dcfl. By Theorem 3.1 and its complement, there exists a finite subset F of $(X \cup \{\#, \$\})^*$ and a basic, confluent, strict, finite c-system S_1 over $(X \cup \{\#, \$\})^*$ such that $\# L\$ = [F]_{\leftrightarrow_{S_1}}$.

We shall now exhibit a left context-sensitive grammar $G = \langle X \cup \{\#, \$\}, V, P, \{\sigma\}\rangle$ which generates $\# L\$$ and such that P is a left-basic, confluent, v-strict, finite semi-Thue system. (Here X is the set of terminal symbols of G, V is the set of non-terminal symbols, P, the set of rules is a finite subset of $(X \cup V)^* \times (X \cup V)^*$ and $\sigma \in V$; by "left context-sensitive" we mean that every rule of P has the form $\alpha v \to \alpha m$ where $v \in V$, α, $m \in (X \cup \{\#, \$\} \cup V)^*$.)

Let $D_1 = \{R_i \times \{u_i\} \times \{v_i\}\}_{i \in [1, n]}$ be a finite decomposition of S_1. Let $\sim$ be the greatest right-congruence which saturates every R_i ($i \in [1, n]$) and also the language $\mathrm{Irr}(S_1)$. We consider the complete, deterministic, finite automaton A which computes $\sim$: $A = \langle X \cup \{\#, \$\}, Q, q_0, d, Q\rangle$ where $Q = (X \cup \{\#, \$\})^*/\sim$, $q_0 = [\varepsilon]_\sim$, $\forall x \in X \cup \{\#, \$\}$, $\forall f \in X \cup \{\#, \$\}^*$, $d([f]_\sim, x) = [fx]_\sim$.

Let $Q_i = \{q \in Q, q \subset R_i\}$ (hence $R_i = \bigcup_{q \in Q_i} q$) and $Q_0 = \{q \in Q, q \subset \mathrm{Irr}(S_1)\}$ (hence $\mathrm{Irr}(S_1) = \bigcup_{q \in Q_0} q$). As S_1 is strongly injective, every R_i ($i \in [1, n]$) is included in $\mathrm{Irr}(S_1)$, hence $\bigcup_{i=1}^n Q_i \subset Q_0$.

We define an alphabet V of non-terminals by

$$V = \bigcup_{i=1}^{n} \{[p, u_i, q] \mid p \in Q_i, q \in Q \text{ and } p.u_i = q\}$$
$$\cup \{[p, x, q] \mid p \in Q_0, q \in Q, x \in X \cup \{\#, \$\}, p.x = q\}$$
$$\cup \{\sigma\}.$$

5.2. Remark. The symbols $[p, u_i, q]$ or $[p, x, q]$ are considered as triples, so that if $u_i = x \in X \cup \{\#, \$\}$, then $[p, u_i, q] = [p, x, q]$ and if $u = u_i = u_j$ and $p \in Q_i \cap Q_j$, then $[p, u_i, q] = [p, u_j, q]$.

5.3. Remark. As some words u_i may be equal to ε, V may contain some symbols of the form $[p, \varepsilon, p]$ for $p \in Q_i$, Let us use the notation

$$\mathrm{Left}(S) = \{u_i\}_{i \in [1,n]}, \qquad \overline{\mathrm{Left}}(S) = \mathrm{Left}(S) - \{\varepsilon\},$$

$$W = V - \{[p, \varepsilon, p] \mid p \in Q_0\} - \{\sigma\}, \qquad X' = X \cup \{\#, \$\}.$$

Definition of P: P is constituted of all the rules of the following types:

(1) $\sigma \to [p_1, Z_1, q_1][p_2, Z_2, q_2] \ldots [p_k, Z_k, q_k] \ldots [p_m, Z_m, q_m]$
where $p_1 = q_0$, $\forall k \in [1, m]$, $[p_k, Z_k, q_k] \in W$, $\forall k \in [1, m-1]$, $q_k = p_{k+1}$, $Z_1 Z_2 \ldots Z_k \ldots Z_m \in F$.

(2) $[p, u_i, q] \to [p_1, Z_1, q_1][p_2, Z_2, q_2] \ldots [p_k, Z_k, q_k] \ldots [p_m, Z_m, q_m]$
where $p_1 = p$, $\forall k \in [1, m]$, $[p_k, Z_k, q_k] \in W$, $\forall k \in [1, m-1]$, $q_k = p_{k+1}$, $Z_1 Z_2 \ldots Z_k \ldots Z_m = v_i$.

(3) $[p, Z, q][p', x, q'] \to [p, Z, q]x$
where $q = p'$, $q \in Q_0$, $x \in X \cup \{\$\}$, $Z \in X \cup \{\#\} \cup \overline{\mathrm{Left}}(S)$.

(4) $[q_0, \#, q] \to \#$.

(5) $[p, Z, q] \to [p, Z, q][q, \varepsilon, q]$
where $Z \in X \cup \{\#\} \cup \overline{\mathrm{Left}}(S)$, $q \in Q_i$ and i is such that $u_i = \varepsilon$.

Let $\varphi : (X' \cup V - \{\sigma\})^* \to X'^*$ be the homomorphism which preserves every letter of X' and removes the brackets and states of the letters of $V - \{\sigma\}$. In other words

$$\text{if } x \in X', \qquad \varphi(x) = x,$$

$$\text{if } [p, Z, q] \text{ is a letter of } V, \quad \varphi([p, Z, q]) = Z.$$

We define the rightmost derivation of G by $f \to_{G^{\mathrm{rm}}} g$ iff $f = \alpha u \beta$, $g = \alpha v \beta$ for some rule $(u, v) \in P$, some word $\alpha \in (X' \cup V)^*$ and some right-context $\beta \in (X')^*$. The relations $\xrightarrow{*}_{G^{\mathrm{rm}}}$ (rightmost derivation) and $\xrightarrow{+}_{G^{\mathrm{rm}}}$ are then deduced of $\to_{G^{\mathrm{rm}}}$ as usual. We note $f \vdash_{G^{\mathrm{rm}}} g$ iff $g \to_{G^{\mathrm{rm}}} f$.

5.4. Lemma. *Let $f \in (X' \cup V - \{\sigma\})^*$. $\sigma \xrightarrow{+}_{G^{\mathrm{rm}}} f$ iff the following six conditions are verified*:

(i) $\varphi(f) \in \#L\$$;

(ii) $f \in V^* X'^*$;

(iii) *if $[p, Z, q][p', Z, q']$ is a factor of f, then $q = p'$*;

(iv) *if $[p, Z, q][p', Z, q']$ is a factor of f, then $q \in Q_0$*;

(v) *if $f \notin X'^*$, then its first letter has the form $[q_0, Z, q]$*;

(vi) *if $[p, \varepsilon, p]$ is a letter of f, then $f \in W^+[p, \varepsilon, p]X'^*$. In other words, the only possible occurrence of a letter $[p, \varepsilon, p]$ is in the rightmost position of the block of non-terminals, and this position must be on the right of at least one non-terminal. In the above conditions, Z denotes any element of* $\mathrm{Left}(S) \cup X'$.

Proof. (1) If $\sigma \rightarrow^{1}_{G^{\mathrm{rm}}} f$, f is the right-hand side of a rule of type (1), hence conditions (i)-(iv) are verified. Let us suppose that $\sigma \rightarrow^{n}_{G^{\mathrm{rm}}} g \rightarrow^{1}_{G^{\mathrm{rm}}} f$ for some $n \in \mathbb{N}-\{0\}$.

Case 1: The last step, $g \rightarrow^{1}_{G^{\mathrm{rm}}} f$, uses a rule of type (2). Then

$$g=[p'_1, Z'_1, q'_1]\ldots[p'_s, Z'_s, q'_s][p, u_i, q]w \quad \text{where } w \in X'^*$$

$$f=[p'_1, Z'_1, q'_1]\ldots[p'_s, Z'_s, q'_s][p_1, Z_1, q_1]\ldots[p_k, Z_k, q_k]\ldots[p_m, Z_m, q_m]w.$$

By induction hypothesis, conditions (i)-(vi) are verified by g.

- (i) is verified by f, because every rule (r_i, u_i, v_i) (where $r_i \in R_i$) saturates $\#L\$$.
- (ii) is clearly verified.
- (iii) is true for g, hence it is true in the prefix $[p'_1, Z_1, q'_1]\ldots[p'_s, Z_s, q'_s]$, $q'_s=p$ and by the form of rules of type (2) $p=p_1$. Moreover, (iii) is true for every right-hand side of P. Hence (iii) is true for f.
- By induction hypothesis, for every $j \in [1, s]$, $q'_j \in Q_0$. As S_1 is strongly injective, every strict prefix of $r_i v_i$ is S_1-irreducible. Hence, for every $k \in [1, m-1]$, $q_k \in Q_0$. So (iv) is true for f.
- If $g=[p, u_i, q]$, then $p=q_0$ hence $p'_1=q_0$ and (v) is true for f. Otherwise, $p'_1=q_0$ and (v) is true for f.
- By induction hypothesis, for every $j \in [1, s]$, $Z'_j \neq \varepsilon$. By definition of the rules of type (2), no $Z_k (k \in [1, m])$ can be equal to ε. Hence (vi) is true for f.

Case 2: The last step, $g \rightarrow^{1}_{G^{\mathrm{rm}}} f$, uses a rule of type (3) or (4). Then conditions (i)-(vi) for f can be easily deduced from the same conditions for g.

Case 3: The last step $g \rightarrow^{1}_{G^{\mathrm{rm}}} f$, uses a rule of type (5). Then

$$g=[p'_1, Z'_1, q'_1]\ldots[p'_s, Z'_s, q'_s][p, Z, q]w \quad \text{where } w \in X'^*$$

$$f=[p'_1, Z'_1, q'_1]\ldots[p'_s, Z'_s, q'_s][q, \varepsilon, q]w.$$

Using the induction hypothesis, we see that conditions (i), (ii), (iii), (v) are true for f. By definition of the rules of type (5), $q \in \bigcup_{i=1}^{n} Q_i$ hence $q \in Q_0$. So (iv) is true for f. By the induction hypothesis, for every $j \in [1, s]$, $Z'_j \neq \varepsilon$. By definition of the rules of type (5), $Z \neq \varepsilon$. Hence condition (vi) is true for f.

We have proved that every right sentential form of G fulfils conditions (i)-(vi).

(2) Let us prove the converse statement. We consider the following valuation $\|*\|$ over $(X' \cup V)^*$:

$$\forall x \in X', \quad \|x\|=2, \qquad \forall v \in V, \quad \|v\|=1.$$

Let us show by induction on $\|f\|$ that, if (i)-(vi) are true for f, then $\sigma \xrightarrow{+}_{G^{\mathrm{rm}}} f$. Let us first observe that the relation $\vdash_{G^{\mathrm{rm}}}$ (as well as $\rightarrow_{G^{\mathrm{rm}}}$) preserves the conjunction of conditions (i)-(vi). To be precise if $f \vdash_{G^{\mathrm{rm}}} g$ and conditions (i)-(vi) are true for f, either $g=\sigma$ or conditions (i)-(vi) are true for g (this can be proved by arguments similar to those used in part (1) of this proof). Then, it is enough to prove that every word f such that (i)-(vi) are true for f, is $\vdash_{G^{\mathrm{rm}}}$-reducible.

Let us consider some $f \in (X' \cup V - \{\sigma\})^*$ such that (i)-(vi) are true for f. By condition (ii), $f = [p_1, Z_1, q_1] \dots [p_j, Z_j, q_j] \dots [p_s, Z_s, q_s]w$ where $w \in X'^*$. By condition (i), $\varphi(f) \in \# L\$$ and $\varphi(f) = Z_1 \dots Z_j \dots Z_s w$.

Case 1: $\varphi(f)$ is S_1-irreducible. Hence $\varphi(f) \in F$.

Subcase 1: $w = \varepsilon$. Here, f is the right-hand side of a rule of type (1), hence f is $\vdash_{G^{\mathrm{rm}}}$-reducible.

Subcase 2: $w \neq \varepsilon$ and $s \neq 0$. If $Z_s = \varepsilon$, then by condition (vi), $s \geq 2$ and $Z_{s-1} \neq \varepsilon$. Setting $g = [p_1, Z_1, q_1] \dots [p_{s-1}, Z_{s-1}, q_{s-1}]w$ we $g \to^1_{G^{\mathrm{rm}}} f$ (using the rule $[p_{s-1}, Z_{s-1}, q_{s-1}] \to [p_{s-1}, Z_{s-1}, q_{s-1}][q_{s-1}, \varepsilon, q_{s-1}]$). If $Z_s \neq \varepsilon$, then f is $\vdash_{G^{\mathrm{rm}}}$-reducible by a rule of type (3).

Subcase 3: $w \neq \varepsilon$ and $s = 0$. Here, $f = \# w'$ where $w' \in X'^*$ hence f is $\vdash_{G^{\mathrm{rm}}}$-reducible by rule (4).

Case 2: $\varphi(f)$ is S_1-reducible. So, $\varphi(f) = r_i v_i w'$ where $i \in [1, n]$, (r_i, u_i, v_i) is a rule of the component (R_i, u_i, v_i) and $w' \in X'^*$.

Subcase 1: $|Z_1 \dots Z_j \dots Z_s| < |r_i v_i|$. As S_1 is strongly injective, $Z_1 \dots Z_j \dots Z_s$ is S_1-irreducible. By conditions (iii) and (v), we then deduce that $q_s \in Q_0$. If $Z_s = \varepsilon$, then f is $\vdash_{G^{\mathrm{rm}}}$-reducible (by a rule of type (5)), otherwise f is $\vdash_{G^{\mathrm{rm}}}$-reducible by a rule of type (4) or (3).

Subcase 2: $|r_i v_i| \leq |Z_1 \dots Z_j \dots Z_s|$. By definition of the alphabet V and by conditions (iii) and (v), for every $j \in [1, s]$, if $|Z_j| \geq 2$, then there exists some $i_j \in [1, n]$ such that $Z_j = u_{i_j}$ and $Z_1 \dots Z_{j-1} \in R_{i_j}$. As S_1 is basic, there is no overlap between v_i and any word Z_j ($j \in [1, s]$). Hence there exists two integers k, l such that $0 \leq k < l \leq s$, $Z_1 \dots Z_k = r_i$ and $Z_{k+1} \dots Z_l = v_i$. Moreover $l = s$ because by condition (iv) $q_{s-1} \in Q_0$ and $Z_1 \dots Z_l$ is S_1-reducible. Let us consider the word $\beta = [p_{k+1}, Z_{k+1}, q_{k+1}] \dots [p_l, Z_l, q_l]$. $p_{k+1} = q_k \in Q_i$; $Z_l \neq \varepsilon$ because $q_l \notin Q_0$; for every $j \in [k+1, l-1]$, $Z_j \neq \varepsilon$ by condition (vi), $\forall j \in [k+1, l-1]$, $q_j = p_{j+1}$; $Z_{k+1} \dots Z_l = v_i$. Hence β is the right-hand side of a rule of type (2). As $l = s$, all the letters which are on the right of the given occurrence of β are in X'. Hence f is $\vdash_{G^{\mathrm{rm}}}$-reducible. □

5.5. Lemma. *If $\sigma \xrightarrow{*}_G \alpha[p, Z, q]\beta$ for some $\alpha, \beta \in (X' \cup V)^*$ and $[p, Z, q] \in V$, then $q_0 . \theta_{S_1}(\varphi(\alpha)) = p$.*

Proof. We prove this lemma by induction on the integer n such that $\sigma \to^n_G \alpha[p, Z, q]\beta$. If $n = 1$, the definition of the rules of type (1) shows that $q_0 . \theta_{S_1}(\varphi(\alpha)) = q_0 . \varphi(\alpha) = p$. Let us suppose that $\sigma \to^n_G g \to^1_G \alpha[p, Z, q]\beta = f$ for some $n \geq 1$. We make the induction hypothesis that for every decomposition $g = \alpha'[p', Z', q']\beta'$ it is true that $q_0 . \theta_{S_1}(\varphi(\alpha')) = p'$.

If the rule used in the last step of this derivation is of type (3) or (4), it is straightforward to show that $q_0 . \theta_{S_1}(\varphi(\alpha)) = p$. If the rule used in the last step of this derivation is of type (5), then

$$g = \alpha'[p', Z', q']\beta', \qquad f = \alpha'[p', Z', q'][q', \varepsilon, q']\beta'$$

where $[p', Z', q'] \to [p', Z', q'][q', \varepsilon, q']$ is a rule of P.

If $|\alpha|\leq|\alpha'|$, $q_0.\theta_{S_1}(\varphi(\alpha))=p$ by induction hypothesis.

If $|\alpha|=|\alpha'|+1$, then $\alpha=\alpha'[p',Z',q']$ and $[p,Z,q]=[q',\varepsilon,q']$. $q_0.(\theta_{S_1}(\varphi(\alpha'))Z')=p'.Z'=q'$. As $q'\in\bigcup_{i=1}^{n} Q_i$, $\theta_{S_1}(\varphi(\alpha'))Z'$ is S_1-irreducible, hence $\theta_{S_1}(\varphi(\alpha')Z')=\theta_{S_1}(\varphi(\alpha'))Z'$. As $\varphi(\alpha)=\varphi(\alpha')Z'$ we can conclude that $q_0.\theta_{S_1}(\varphi(\alpha))=q'=q$. which is the required property.

If $|\alpha|\geq|\alpha'|+2$, then $\alpha=\alpha'[p',Z',q'][q',\varepsilon,q']\beta_1'$ and $\beta'=\beta_1'\beta_2'$ for some β_1', $\beta_2'\in(X'\cup V)^*$. But $\varphi(\alpha)=\varphi(\alpha'[p',Z',q']\beta_1')$. Hence, by induction hypothesis $q_0.\theta_{S_1}(\varphi(\alpha))=p$.

If the rule used in the last step of the derivation is of type (2) then

$$g=\gamma[r,u_i,r']\delta \qquad \text{where } i\in[1,n],\ \gamma,\delta\in(X'\cup V)^*, [r,u_i,r']\in V;$$

$$f=\gamma w\delta \quad \text{and} \quad w=[p_1,Z_1,q_1]\ldots[p_k,Z_k,q_k]\ldots[p_m,Z_m,q_m]$$

where

$$p_1=r, \qquad \forall k\in[1,m],\quad [p_k,Z_k,q_k]\in W$$

$$\forall k\in[1,m-1],\quad q_k=p_{k+1},\quad Z_1\ldots Z_k\ldots Z_m=v_i.$$

We distinguish three cases.

Case 1: $|\alpha|<|\gamma|$. In other words, the given occurrence of $[p,Z,q]$ is in γ. Hence, by the induction hypothesis, $q_0.\theta_{S_1}(\varphi(\alpha))=p$.

Case 2: $|\gamma|\leq|\alpha|<|\gamma w|$. In other words, the given occurrence of $[p,Z,q]$ is the kth letter of w, for some $k\in[1,m]$. By the induction hypothesis, $q_0.\theta_{S_1}(\varphi(\gamma))=r$. By the definition of rules of type (2), $r=p_1$, $p_1.(Z_1\ldots Z_{k-1})=q_{k-1}$ and $q_{k-1}=p_k$. As $q_{k-1}\in Q_0$ (because S_1 is strongly injective) $\theta_{S_1}(\varphi(\alpha))=\theta_{S_1}(\varphi(\gamma)).\ Z_1\ldots Z_{k-1}$. Hence

$$q_0.\theta_{S_1}(\varphi(\alpha))=[q_0.\theta_{S_1}(\varphi(\gamma))].Z_1\ldots Z_{k-1}=r.Z_1\ldots Z_{k-1}=q_{k-1}=p.$$

Case 3: $|\gamma w|\leq|\alpha|$. Hence $\delta=\delta'\delta''$ for some δ', $\delta''\in(X'\cup V)^*$ and $\alpha=\gamma w\delta'$. By induction hypothesis,

$$q_0.\theta_{S_1}(\varphi(\gamma[r,u_i,r']\delta'))=p \quad \text{and} \quad q_0.\theta_{S_1}(\varphi(\gamma))=r.$$

As $r\in Q_i$, $\varphi(\gamma)u_i \overset{*}{\leftrightarrow}_{S_1} \varphi(\gamma)v_i$. Hence

$$\theta_{S_1}(\varphi(\gamma[r,u_i,r']\delta'))=\theta_{S_1}(\varphi(\gamma w\delta'))$$

so that $q_0.\theta_{S_1}(\varphi(\gamma w\delta'))=p$. $\square$

5.6. Lemma. *Let $f\in X'^*$. $\sigma\xrightarrow{*}_G f$ iff $f\in\#L\$$. In other words, the language generated by G is $\#L\$$.*

Proof. Let $f\in\#L\$$. Then f fulfils conditions (i)–(vi) of Lemma 5.4. Hence $\sigma\xrightarrow{+}_{G^{\mathrm{rm}}} f$, so $\sigma\xrightarrow{*}_G f$. Let us prove by induction on the integer $n\geq1$ that, for every $g\in(X'\cup V-\{\sigma\})^*$ if $\sigma\rightarrow^n_G g$ then $\varphi(g)\in\#L\$$. If $n=1$, the property is true. Let us suppose that $\sigma\rightarrow^n_G h\rightarrow^1_G g$.

If the rule used in the last step is not of type (2), then $\varphi(g)=\varphi(h)$ and by induction hypothesis, $\varphi(h)\in \#L\$$. If the rule used in the last step is of type (2) then $h=\alpha[p, u_i, q]\beta$; $g=\alpha w\beta$ for some $\alpha, \beta\in(X'\cup V)^*$ and some rule of type (2), $[p, u_i, q]\to w$, where $\varphi(w)=v_i$. By Lemma 5.5 $q_0.\theta_{S_1}(\varphi(\alpha))=p$, hence $\theta_{S_1}(\varphi(\alpha))\in R_i$. So $\varphi(h) \overset{*}{\leftrightarrow}_{S_1} \varphi(g)$. By induction hypothesis $\varphi(h)\in \#L\$$ which is saturated by $\overset{*}{\leftrightarrow}_{S_1}$, hence $\varphi(g)\in \#L\$$. □

Let us see now that the semi-Thue system P is confluent . Given a semi-Thue system S over an alphabet Y, we say that a pair of rules of S, $((u, v), (u', v'))$, is overlapping, iff either

(I) $va=bv'$ for some $a, b\in Y^+$ such that $|b|<|v|$, or
(II) $v=av'b$ for some $a, b\in Y^*$ and $(u, v)\neq(u', v')$.

5.7. Remark. The property to overlap is defined as a property of the ordered pair $((u, v), (u', v'))$ and not of the set $\{(u, v), (u', v')\}$. It might happen that a pair $((u, v), (u, v))$ is overlapping.

We call a semi-Thue system *overlap-free* iff it has no overlapping pair of rules. From the well known confluence criterium (see [21]) it is clear that every overlap-free semi-Thue system is confluent.

In order to have a system P which is overlap-free, we must choose a c-system S_1 which is minimal with respect to $\#L\$$ in the sense detailed below.

Let S be a c-system and L a language over a finite alphabet X, such that $\leftrightarrow_S$ saturates L. S is said to be *L-minimal* iff, for every rule $(r, u, v)\in S$, there exists $w\in X^*$ such that $ruw\in L$.

In the next lemma we show that in Theorem 3.1, the c-system S can be chosen L-minimal.

5.8. Lemma. *Let X be a finite alphabet. Let $L=[R]_{\leftrightarrow_S}$ where R is a rational language over X and S is a basic, strongly injective, strict, finite* c*-system. Then there exists a* c*-system $S_1\subset S$ such that $L=[R]_{\leftrightarrow_{S_1}}$, S_1 is also basic, strongly injective, strict and finite and, in addition, S_1 is L-minimal.*

Proof. Let $D=\{R_i\times\{u_i\}\times\{v_i\}\}_{i\in[1,n]}$ be a finite decomposition of a c-system S over X, which is basic, strongly injective, strict and finite. Let R be a rational set over X. For every $i\in[1, n]$, we define $R_i'=\{r\in R_i, \exists w\in X^*, ru_iw\in L\}$. Let us set $S_1=\bigcup_{i=1}^n R_i'\times\{u_i\}\times\{v_i\}$. It is easy to check that $L=[R]_{\leftrightarrow_{S_1}}$ and S_1 is still basic, strongly injective and strict. To prove Lemma 5.8 it suffices to show that every R_1' is rational. Let us consider the set $H=L(X^*)^{-1}\cap \mathrm{Irr}(S)$. By Lemma 4.20, H is rational. We claim now that

$$\forall i\in[1, n],\quad R_i'=R_i\cap(\langle H\rangle_{\leftarrow_S}u_i^{-1}) \qquad (\mathscr{E}_i)$$

(1) Let $r\in R_i'$. Then, there exists $w\in X^*$ such that $ru_iw\in L$. Hence $\theta_S(ru_i)w\in L$, so $\theta_S(ru_i)\in H$. The fact that S is strongly injective and left-basic implies that the reduction $ru_i \overset{*}{\vdash}_S \theta_S(ru_i)$ is right-linear. Hence $r\in R_i\cap(\langle H\rangle_{\leftarrow_S}u_i^{-1})$.

(2) Let $r \in R_i \cap (\langle H \rangle_{\xrightarrow{*}_{S^r}} u_i^{-1})$. There exists $h \in H$ such that $ru_i \vdash^*_{S^r} h$ and $r \in R_i$. By definition of H, there exists $w \in X^*$ such that $hw \in L$. As $\vdash^*_{S^r}$ saturates L, $ru_iw \in L$, hence $r \in R'_i$. Hence every equality $(\mathscr{E}_i)$ is true. By Lemma 3.6 the right-hand side of $\mathscr{E}_i$ is rational, hence R'_i is rational. □

We can go back to our semi-Thue system P. From now on, we suppose that the c-system S_1 from which we have built P is $\#L\$$-minimal. By Lemma 5.8 this assumption can be made. We also suppose that D_1 is such that for every $i \in [1, n]$, $R_i \neq \emptyset$.

5.9. Lemma. *P is overlap-free.*

Proof. In part (1) we investigate all the pairs of P and show that an overlap of type I is impossible. In part (2) we do the same work for overlaps of type II.

(1) For every pair $(i, j) \in [1, 5] \times [1, 5]$, with $i \leq j$, we show that a pair $((u, v), (u', v'))$, where (u, v) is of type (i) and (u', v') of type (j), or where (u', v') is of type (i) and (u, v) of type (j), cannot be overlapping. We have omitted some pairs (i, j) for which the result is trivial.

(1, 1): Impossible, because in a rule of type (1), $Z_1 \in \#X^*$, $Z_m \in X^*\$$ and the other symbols Z_k belong to X^*.

(1, 2): As S_1 is minimal, if $v' = [p'_1, Z'_1, q'_1] \dots [p'_{m'}, Z'_{m'}, q'_{m'}]$, the only possible occurrence of $\#$ (resp. \$) is in Z'_1 (resp. $Z'_{m'}$). The above argument can then be applied.

(1, 3): Impossible because $v = \gamma[p_m, Z_m, q_m]$ and $v' = [p, Z, q]x$ where $Z_m \in X^*\$$ while $Z \notin X^*\$$ and $\gamma \in V^*$ while $x \notin V$.

(1, 5): Impossible because $v = [p_1, Z_1, q_1]\gamma[p_m, Z_m, q_m]$ and $v' = [p, Z, q][q, \varepsilon, q]$ where $Z_1 \in \#X^*$ (hence $Z_1 \neq \varepsilon$) and $Z_m \in X^*\$$ (hence $Z_m \neq Z$).

(2, 2): Impossible because $v = [p_1, Z_1, q_1] \dots [p_m, Z_m, q_m]$, $v' = [p'_1, Z'_1, q'_1] \dots [p'_{m'}, Z'_{m'}, q'_{m'}]$. But q_m, $q'_{m'} \in Q - Q_0$ while for every $k \in [1, m-1]$ (resp. $k' \in [1, m'-1]$) the fact that S_1 is strongly injective implies that q_k (resp. $q'_{k'}$) $\in Q_0$.

(2, 3): Impossible because $v = [p_1, Z_1, q_1] \dots [p_m, Z_m, q_m]$, $v' = [p, Z, q]x$ where $x \notin W$ and $q \in Q_0$ while $q_m \notin Q_0$.

(2, 5): Impossible because $v = [p_1, Z_1, q_1] \dots [p_m, Z_m, q_m]$, $v' = [p, Z, q][q, \varepsilon, q]$ where $[q, \varepsilon, q] \notin W$ and $q_0 \in Q_0$ while $q_m \notin Q_0$.

(3, 5): Impossible because $v = [p, Z, q]x$ where $Z \neq \varepsilon$ and $v' = [p', Z', q'][q', \varepsilon, q']$ where $Z' \neq \varepsilon$.

(2) Here we consider all the pairs $(i, j) \in [1, 5] \times [1, 5]$ except those for which it is clear that an overlap of type II cannot occur.

(1, 1): Impossible because of the special letters $\#$, \$.

(1, 2): Impossible because $v = [p_1, Z_1, q_1] \dots [p_m, Z_m, q_m]$, $v' = [p'_1, Z'_1, q'_1] \dots [p'_{m'}, Z'_{m'}, q'_{m'}]$ where $q'_{m'} \notin Q_0$ while every q_k $(k \in [1, m])$ is in Q_0.

(1, 3); Impossible because $v' \notin V^*$ while $v \in V^*$.

(1, 4): Idem.

(1, 5): Impossible because $[q, \varepsilon, q] \notin W$ while $v \in W^*$.

(2, 1): Impossible because $v = [p_1, Z_1, q_1] \ldots [p_m, Z_m, q_m]$, $v' = [p'_1, Z'_1, q'_1] \ldots [p'_{m'}, Z'_{m'}, q'_{m'}]$ where $Z'_1 \in \#X^*$ and $Z'_{m'} \in X^*\$$ (because $F \subset \#L\$ \subset \#X^*\$$). As S_1 is $\#L\$$-minimal, if v' is factor of v then we must have $Z'_1 = Z_1$ and $Z'_{m'} = Z_m$ hence $v = v'$. Let $r \in R_i$, where $[p_1, u_i, q_1]$ is the left-hand side of (u, v). As $Z_1 \in \#X^*$ and S_1 is $\#L\$$-minimal, we must have $r = \varepsilon$. Hence $Z_1 \ldots Z_m$ is S_1-reducible while $Z'_1 \ldots Z'_m$ is S_1-irreducible, contradicting the equality $v = v'$.

(2, 2): Impossible because S_1 is strongly injective.

(2, 3): Impossible because $v \in V^*$ while $v' \notin V^*$.

(2, 4): Idem.

(2, 5): Impossible because $v \in W^*$ while $v' \notin W^*$.

(3, 1): Impossible because $|v'| \geq 2$ and $v' \in V^*$ while the only factor of v which belongs to V^* has a length equal to one.

(3, 2): Impossible because $v = [p, Z, q]x$, $v' = [p_1, Z_1, q_1] \ldots [p_m, Z_m, q_m]$ and $q \in q_0$ while $q_m \in Q - Q_0$.

(3, 4): Impossible because $\#$ is not a letter occurring in v.

(3, 5): Impossible because $|v| = |v'|$ but the last letter of v', some $[q, \varepsilon, q]$, cannot occur in v.

(5, 1): The argument used in the case (3, 1) also applies here.

(5, 2): Impossible because $v = [p, Z, q][q, \varepsilon, q]$, $v' = [p_1, Z_1, q_1] \ldots [p_m, Z_m, q_m]$ where $q \in \bigcup_{i=1}^{n} Q_i \subset Q_0$ while $q_m \in Q - Q_0$. $\square$

5.10. Lemma. *P is left-basic.*

Proof. One can easily check that no left-hand side of a rule of P can strictly contain a right-hand side of rule. Hence condition (C1) of Definition 2.2 is verified.

As the left-hand sides of rules of type 1, 2, 4, 5 are letters, they cannot be overlapped on the left by some right-hand side of rule. Hence the only violation of condition (C2) which could be possible would consist of a word $u = [p, Z, q][p', x, q']$ (left-hand side of a rule of type (3)) which would be overlapped on the left by some right-hand side of rule. Since $Z \notin X^*\$ \cup \{\varepsilon\}$ and $q \in Q_0$, such an overlap is impossible. $\square$

We can now end the proof of Proposition 5.1. Let $Y = X \cup \{\#, \$\} \cup V$, $f = \sigma$ and $S = P$. By Lemma 5.6, $\#L\$ = \langle f \rangle_{\xrightarrow{*}_S} \cap (X \cup \{\#, \$\})^*$. By Lemma 5.9, S is overlap-free, hence confluent. So

$$\#L\$ = [f]_{\xleftrightarrow{*}_S} \cap (X \cup \{\#, \$\})^*.$$

By Lemma 5.10, S is left-basic and in addition, S is v-strict (with respect to $\|*\|$). Hence Proposition 5.1 is proved.

5.11. Remark. (1) The fact that we start with a c-system S_1 which is basic (and not merely left-basic) is used in the proof of Lemma 5.4 (see part (2) of the proof, Case 2, Subcase 2).

(2) In Proposition 5.1, the property "left-basic" cannot be strengthened into "basic". The reason is that every language of the form $[f]_{\leftrightarrow_S}$ where S is a basic, confluent, v-strict semi-Thue system is a generalised NTS language [33, Chapter VI]. As the family of g.NTS languages is preserved by intersection with rational sets and some marked dcfls $\# L\$$ are not g.NTS languages, they cannot be represented as $[f]_{\leftrightarrow_S} \cap (X \cup \{\#, \$\})^*$ with S basic, confluent and v-strict.

(3) In the proof of Proposition 5.1, we could start with a $LR(0)$ grammar G which generates $\# L\$$. Then, the finite set of triples (R, v, m) where (v, m) is a rule of G and R is the set of left-contexts r such that rv is a prefix of some right-sentential form of G, furnishes a c-system S_2 which is monadic, strongly injective and finite (this is noticed, in a different vocabulary, in [1, p. 397, Exercise 5-2-10]). We can then end the proof with that c-system S_2 as we did with our c-system S_1. Hence our Theorem 3.1 cannot be considered as an obliged preliminary of Proposition 5.1. Nevertheless, we cannot extend this remark to our next results (Theorems 5.12 and 5.17).

We now want to establish the following result, which strengths the conditions that can be imposed on S in Proposition 5.1.

5.12. Theorem. *Let L be a language over the finite alphabet X and let $\#$, $\$$ be new letters. L is deterministic context-free if and only if there exists an alphabet Y containing $X' = X \cup \{\#, \$\}$, a word $f \in Y^*$, a left-basic, confluent,* v-*strict, finite semi-Thue system S over Y, an homomorphism $\varphi : Y^* \to X^*$ and a rational set R over Y, such that*

$$\# L\$ = [f]_{\leftrightarrow_S} \cap X'^*, \qquad [f]_{\leftrightarrow_S} = \varphi^{-1}(\# L\$) \cap R$$

where $X'^* \subset R$ and $\varphi_{|X'^*} = \mathrm{Id}_{X'^*}$.

The only thing that remains to be proved is that, given a dcfl L, one can find S fulfilling the properties stated in Theorem 5.12. Let L be a dcfl over a finite alphabet X. We consider the left-context-sensitive grammar G built in the proof of Proposition 5.1. Let us denote by $\mathrm{RMS}(G)$ the set of rightmost sentential forms of G: $\mathrm{RMS}(G) = \{f \in (X' \cup V)^*, \sigma \xrightarrow{*}_{G^{\mathrm{rm}}} f\}$. Let us denote by R the following set:

$$R = \{r \in (X' \cup V - \{\sigma\})^* \mid r \text{ fulfils all conditions (ii)–(vi) of Lemma 5.4}\} \cup \{\sigma\}.$$

Let us extend the homomorphism φ by setting $\varphi(\sigma) = \# m_0\$$ *where m_0* is some fixed word of L (m_0 can be arbitrarily chosen). Then we have

$$\# L\$ = \mathrm{RMS}(G) \cap X'^*, \tag{$\mathscr{E}_1$}$$

$$\mathrm{RMS}(G) = \varphi^{-1}(\# L\$) \cap R \tag{$\mathscr{E}_2$}$$

where $X'^* \subset R$ and $\varphi|_{X'^*} = \mathrm{Id}_{X'^*}$.

It is sufficient now to prove that $\mathrm{RMS}(G) = [\sigma]_{\leftrightarrow_T}$ for some left-basic, confluent, v-strict, finite semi-Thue system T to prove Theorem 5.12. We consider the following semi-Thue systems:

$$T_1 = \{(u, v) \in (X' \cup V)^* \times (X' \cup V)^* \mid (u, v) \in P \text{ and } \varphi(v) \in X'^*\$\},$$

$$T_2 = \{(ux, vx) \mid (u, v) \in P, \varphi(v) \notin X'^*\$ \text{ and } x \in X \cup \{\$\}\},$$

$$T = T_1 \cup T_2.$$

5.13. Lemma. *T is overlap-free.*

Proof. Using the fact that no letter $x \in X \cup \{\$\}$ can be the first letter of a word v such that $(u, v) \in P$, one can show that every overlap of a pair of rules in T leads to an overlap of a pair of rules in P. As P is overlap-free this is impossible. □

5.14. Lemma. *T is left-basic.*

Proof. Let us consider the left-hand side of a rule (α, β) which is overlapped on the left by the right-hand side of a rule (α', β').

Case 1: $(\alpha, \beta) = (u, v) \in P$ and $(\alpha', \beta') = (u', v') \in P$. This case is impossible because P is left-basic.

Case 2: $(\alpha, \beta) = (u, v) \in P$ and $(\alpha', \beta') = (u'x', v'x')$ where $(u', v') \in P$. This case is impossible because P is left-basic and $x' \in X \cup \{\$\}$ cannot be the first letter of u.

Case 3: $(\alpha, \beta) = (ux, vx)$ where $(u, v) \in P$ and $(\alpha', \beta') = (u', v') \in P$. This case is impossible because P is left-basic and u cannot be a suffix of v': $\varphi(u) \notin X'^*\$$ while $\varphi(v') \in X'^*\$$.

Case 4: $(\alpha, \beta) = (ux, vx)$, $(\alpha', \beta') = (u'x', v'x)$ where (u, v) and (u', v') belong to P. The same argument as in Case 2 applies.

Hence, condition (C2) of Definition 2.2 is fulfilled. Let us check now that condition (C1) is also fulfilled. We consider two rules (α, β), $(\alpha', \beta') \in T$ such that $\alpha = r'\beta's'$ for some r', $s' \in (X' \cup V)^*$. We distinguish the same cases as above. Cases 1, 2 and 4 are impossible. Case 3 is possible only when

$$(\alpha, \beta) = (u\$, v\$)$$

$$(u, v) = ([p, Z, q], [p_1, Z_1, q_1] \dots [p_m, Z_m, q_m]) \quad \text{is a rule of type (2) of } P$$

$$(\alpha', \beta') = ([p, Z, q][p', \$, q'], [p, Z, q]\$) \quad \text{is a rule of type (3) of } P$$

Hence, in this case, $\alpha = [p, Z, q]\$ = \beta'$ and $r' = s' = \varepsilon$, so that (C1) is fulfilled. □

5.15. Lemma. $\mathrm{RMS}(G) = [\sigma]_{\leftrightarrow_T}$.

Proof. We prove that $\rightarrow_T$ and $\vdash_T$ are are preserving $\mathrm{RMS}(G)$ and that every word in $\mathrm{RMS}(G) - \{\sigma\}$ is $\vdash_T$-reducible.

(1) Let $f \in \mathrm{RMS}(G)$ and $g \in (X' \cup V)^*$ such that $f \to_T g$. If $f = \sigma$, it is clear that $g \in \mathrm{RMS}(G)$. Let us suppose that $f \neq \sigma$. Then $f = \gamma\alpha\delta$ and $g = \gamma\beta\delta$ for some γ, $\delta \in (X' \cup V)^*$ and $(\alpha, \beta) \in T$. If $(\alpha, \beta) \in T_1$, then $\varphi(\beta) \in X'^*\$$ and hence $\varphi(\alpha) \in X'^*\$$. By Lemma 5.4, $\varphi(\gamma\alpha)\varphi(\delta) \in \#L\$$, so $\varphi(\delta) = \varepsilon$. By Lemma 5.4, point (vi), either $\delta = \varepsilon$ or $\delta = [p, \varepsilon, p]$. But if $\delta = [p, \varepsilon, p]$, then S_1 has a rule of the form (r, ε, v) with $r \in X'^*\$$, which contradicts the $\#L\$$-minimality of S_1. Hence $\delta = \varepsilon$. So $f \to_P g$ and this derivation is rightmost. Hence $g \in \mathrm{RMS}(G)$. If $(\alpha, \beta) \in T_2$, then $f = \gamma ux\delta$ and $g = \gamma vx\delta$ for some $(u, v) \in P$, $x \in X \cup \{\$\}$ and γ, $\delta \in (X' \cup V)^*$. By Lemma 5.4, point (ii), $\delta \in X'^*$, hence $f \to_{G^{\mathrm{rm}}} g$.

(2) By similar arguments, one can show that $\vdash_T$ saturates $\mathrm{RMS}(G)$.

(3) Let $f \in \mathrm{RMS}(G) - \{\sigma\}$. Then there exist $g \in \mathrm{RMS}(G)$, γ, $\delta \in (X' \cup V)^*$ and $(u, v) \in P$ such that $f = \gamma v\delta$, $g = \gamma u\delta$ and $\delta \in X'^*$. If $|\delta| = 0$, as $\varphi(f) \in \#L\$$, we must have $\varphi(v) \in X'^*\$$. Hence $(u, v) \in T_1$ and f is T_1-reducible. If $|\delta| \geq 1$, as $\varphi(f) \in \#L\$$, $\delta = \varphi(\delta) \in X'^*\$$. Hence $\varphi(v) \notin X'^*\$$. Let $\delta = x\delta'$ where $x \in X'$, $\delta' \in X'^*$. Then $(ux, vx) \in T_2$ and f is T_2-reducible.

Lemma 5.15 follows from these three properties. □

We can now end the proof of Theorem 5.12. Let us take $S = T$. S is v-strict with respect to the valuation $\|*\|$ considered in the proof of Lemma 5.4. Then the equalities $\mathscr{E}_1$, $\mathscr{E}_2$ and Lemmas 5.13, 5.14 and 5.15 show that S, R and φ are fulfilling the properties announced in Theorem 5.12.

5.16. Example. Let us compute the semi-Thue systems P and T in the following example.

$$X = \{a, b\}, \qquad L = \{a^{2n}b^{2n}\}_{n \geq 1} \cup \{a^{2n+1}b^{4n+2}\}_{n \geq 0}$$

$$S_1 = \#(a^2)^* \times \{b\} \times \{a^2b^3\} \cup \#(a^2)^*a \times \{b\} \times \{a^2b^5\}$$

$$F = \{\#a^2b^2\$, \#ab^2\$\} \quad \text{and} \quad \#L\$ = [F]_{\leftrightarrow_{S_1}}$$

$$R_1 = \#(a^2)^*, \qquad R_2 = \#(a^2)^*a,$$

$$\mathrm{Irr}(S_1) = \mathrm{C}[[\#(a^2)^+b^3 \cup \#(a^2)^+ab^5]X'^*].$$

The automaton A can be drawn as shown in Fig. 9 with

$$Q = [0, 12] \quad \text{(the set of all integers } n \text{ such that } 0 \leq n \leq 12)$$

$$Q_0 = Q - \{9\}, \qquad Q_1 = \{1, 3\}, \qquad Q_2 = \{2, 4\}.$$

The semi-Thue system P is constituted of the following rules.

(1) $\sigma \to [0, \#, 1][1, a, 2][2, a, 3][3, b, 10][10, b, 11][11, \$, 12]$,

$\sigma \to [0, \#, 1][1, a, 2][2, b, 12][12, b, 12][12, \$, 12]$.

(2) $[1, b, 12] \to [1, a, 2][2, a, 3][3, b, 10][10, b, 11][11, b, 9]$,

$[3, b, 10] \to [3, a, 4][4, a, 3][3, b, 10][10, b, 11][11, b, 9]$,

$[2, b, 12] \to [2, a, 3][3, a, 4][4, b, 5][5, b, 6][6, b, 7][7, b, 8][8, b, 9]$,

$[4, b, 5] \to [4, a, 3][3, a, 4][4, b, 5][5, b, 6][6, b, 7][7, b, 8][8, b, 9]$.

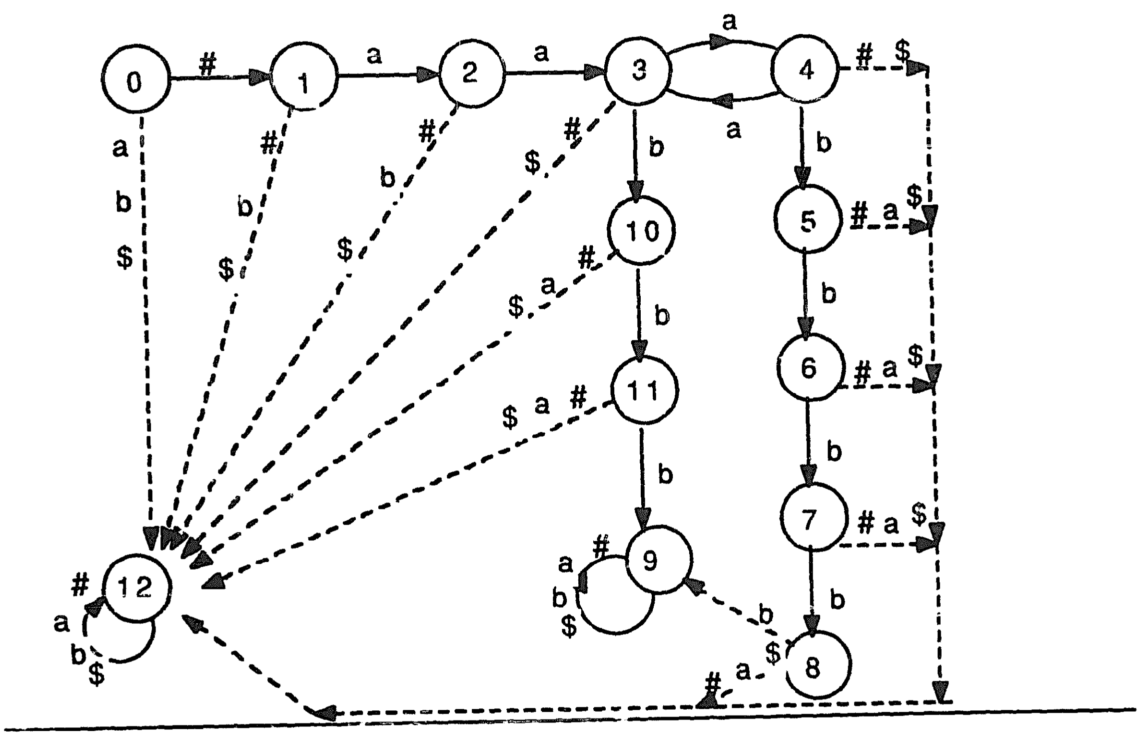

Fig. 9.

(3) $[p, Z, q][q, x, q'] \to [p, Z, q]x$. For every $Z \in \{a, b, \#\}$, $x \in \{a, b, \$\}$, p, q, $q' \in [0, 12]$ such that $q \neq 9$, $p.Z = q$, $q.x = q'$.

(4) $[0, \#, 1] \to \#$. The semi-Thue system T is constituted of the rules (we denote by $(i.j)$ the set of rules of T_i deduced from the rules of type j of P):

(1.1) $\sigma \to [0, \#, 1][1, a, 2][2, a, 3][3, b, 10][10, b, 11][11, \$, 12]$,

$\sigma \to [0, \#, 1][1, a, 2][2, a, 12][12, b, 12][12, \$, 12]$.

(1.3) $[p, Z, q][q, \$, q'] \to [p, Z, q]\$$ for every $Z \in \{a, b, \#\}$, p, q, $q' \in [0, 12]$ such that $q \neq 9$, $p.Z = q$, $q.\$ = q'$.

(2.2) $[1, b, 12]x \to [1, a, 2][2, a, 3][3, b, 10][10, b, 11][11, b, 9]x$,

$[3, b, 10]x \to [3, a, 4][4, a, 3][3, b, 10][10, b, 11][11, b, 9]x$,

$[2, b, 12]x \to [2, a, 3][3, a, 4][4, b, 5][5, b, 6][6, b, 7][7, b, 8][8, b, 9]x$,

$[4, b, 5]x \to [4, a, 3][3, a, 4][4, b, 5][5, b, 6][6, b, 7][7, b, 8][8, b, 9]x$

for every $x \in \{a, b, \$\}$.

(2.3) $[p, Z, q][q, x, q']y \to [p, Z, q]xy$ for every $Z \in \{a, b, \#\}$, $x \in \{a, b\}$, $y \in \{a, b, \$\}$, p, q, $q' \in [0, 12]$ such that $q \neq 9$, $p.Z = q$, $q.x = q'$.

(2.4) $[0, \#, 1]x \to \#x$ for every $x \in \{a, b, \$\}$.

5.2. *Some decision problems about finite semi-Thue systems*

Let us recall that the equivalence problem for dpda is the following problem:

Instance: two dpdas A_1, A_2

Question: Are the languages $L(A_1)$, $L(A_2)$ equal?

The other problems that we consider here involve the notions of syntactic congruence (which is classical) and the notion of partial confluence (which we define below).

Let L be a language over a finite alphabet X. *The syntactic congruence* of L, noted $\equiv_L$, is defined as the greatest congruence $\sim$ (for the inclusion ordering) over the monoid X^* which saturates L (i.e. such that for every $f, g \in X^*$, $f \in L$ and $f \sim g$ imply $g \in L$) [19]. It can be characterised by $f \equiv_L g$ iff for every $\alpha, \beta \in X^*$, $\alpha f \beta \in L \Leftrightarrow \alpha g \beta \in L$. Let S be a finite semi-Thue system over X and let f (resp. K) be a word (resp. a subset) of X^*. We say that S is *partially confluent* on f (resp. on K) iff $\langle f \rangle_{\xrightarrow{*}_S} = [f]_{\xleftrightarrow{*}_S}$ (resp. $\langle K \rangle_{\xrightarrow{*}_S} = [K]_{\xleftrightarrow{*}_S}$).

We shall consider between two problems the relation of *recursive reduction* and *recursive equivalence.* We say that a problem P recursively reduces to a problem P' iff there exists a recursive function φ, which maps every instance x of P into an instance $\varphi(x)$ of P', such that the solution of problem P on x is the same as the solution of problem P' on $\varphi(x)$. P and P' are said to be recursively equivalent iff each one of these problems recursively reduces to the other (we shall sometimes omit the words "recursive" or "recursively" when no confusion is possible). The main result of this section is the following.

5.17. Theorem. *The following four problems are recursively equivalent*:

(1) *the equivalence problem for* dpda;

(2) *the class equivalence problem for left-basic, confluent, strict, finite semi-Thue systems*;

(3) *the word problem for the syntactic congruence of one class specified by a left-basic, confluent, strict, finite semi-Thue system*;

(4) *the problem of partial confluence on a word for left-basic, strict, finite semi-Thue systems.*

Let us define more precisely Problems 2, 3 and 4.

Problem 2 (*Instance*): One word f and two semi-Thue systems S_1, S_2 over an alphabet X, such that S_1, S_2 are left-basic, confluent, strict and finite.

(*Question*): $[f]_{\xleftrightarrow{*}_{S_1}} = [f]_{\xleftrightarrow{*}_{S_2}}$?

(This problem has been investigated in Section 4 for some other classes of c-systems or semi-Thue systems).

Problem 3 (*Instance*): Three words f, g, h and one semi-Thue system S over an alphabet X, such that S is left-basic, confluent, strict and finite.

(*Question*): $f \equiv_L g$ where $L = [h]_{\xleftrightarrow{*}_S}$?

Problem 4 (*Instance*): One word f and one semi-Thue system S over an alphabet X, such that S is left-basic, strict and finite.

(*Question*): $\langle f\rangle_{\xrightarrow{*}_S}=[f]_{\xleftrightarrow{*}_S}$?

The next three propositions will allow us to reduce Problem 1 to Problem 3. The constructions that we make in the proof of these propositions will be also used to reduce Problem 1 to Problem 2 and Problem 1 to Problem 4.

In the following, by *proper language* we mean a language which does not contain the empty word and by *proper semi-Thue system* we mean a semi-Thue system such that, for every rule (u, v), u is not the empty word.

5.18. Proposition. *The equivalence problem for two* dpdas *reduces to the equivalence problem for one* dpda *and one class specified by a left-basic, overlap-free, strict, finite and proper semi-Thue system.*

Proof. Let A_1, A_2 be two dpdas on the terminal alphabet X. Let $L_1=L(A_1)$ and $L_2=L(A_2)$.

By Theorem 5.12, one can find an alphabet Y_1 containing X and two special letters $\#$, $\$$, a word $f_1\in Y_1^*$, a left-basic, confluent, v-strict, finite semi-Thue system S_1 over Y_1^*, an homomorphism $\varphi_1: Y_1^*\to X'^*$ (where X' means $X\cup\{\#,\$\}$) and a rational set R_1 over Y_1 such that

$$\#L_1\$=[f_1]_{\xleftrightarrow{*}_{S_1}}\cap X'^*,\qquad [f_1]_{\xleftrightarrow{*}_{S_1}}=\varphi_1^{-1}(\#L_1\$)\cap R_1,$$

$$X'^*\subset R_1,\ \varphi_{1|X'^*}=\mathrm{Id}_{X'^*}.$$

Moreover, the proof of Theorem 5.12 shows that S_1 can be chosen overlap-free and proper. We claim that $L_1=L_2$ iff $[f_1]_{\xleftrightarrow{*}_{S_1}}=\varphi_1^{-1}(\#L_2\$)\cap R_1$.

(1) Let us suppose that $L_1=L_2$. Then

$$\varphi_1^{-1}(\#L_1\$)\cap R_1=\varphi_1^{-1}(\#L_2\$)\cap R_2,$$

and hence $[f_1]_{\xleftrightarrow{*}_{S_1}}=\varphi_1^{-1}(\#L_2\$)\cap R_2$.

(2) Let us suppose that $[f_1]_{\xleftrightarrow{*}_{S_1}}=\varphi_1^{-1}(\#L_2\$)\cap R_1$, hence

$$\varphi_1^{-1}(\#L_1\$)\cap R_1=\varphi_1^{-1}(\#L_2\$)\cap R_1. \qquad (\mathscr{E})$$

But, since φ_1 restricted to X'^* is the identity mapping in X'^*, and X'^* is included in R_1, for every language $L\subset X'^*$, $\varphi_1(\varphi_1^{-1}(L)\cap R_1)=L$. Applying φ_1 to both sides of equality ($\mathscr{E}$), we obtain $\#L_1\$=\#L_2\$$, hence $L_1=L_2$. Our claim is then proved.

In order to deal with a strict semi-Thue system (instead of a merely v-strict semi-Thue system) we use the following trick. Let $m=\max_{y\in Y_1}(\nu(y))$ (where ν is a valuation such that S_1 is v-strict with respect to ν). For every $i\in[1, m]$ we define a copy Y_i of Y_1; all copies are disjoint and every $y\in Y_1$ has a copy y_i in Y_i ($i\in[1, m]$, in the case where $i=1$, $y_1=y$); we define $\hat{Y}=\bigcup_{i=1}^m Y_i$ and $\psi: Y_1^*\to \hat{Y}^*$ is the homomorphism defined by $\forall y\in Y_1$, $\psi(y)=y_1y_2\ldots y_p$ with $p=\nu(y)$. We then define $S=\psi(S_1)=\{(\psi(u), \psi(v))\mid (u, v)\in S_1\}$, as for every $w\in Y_1^*$, $|\psi(w)|=\nu(w)$, $\psi(S_1)$ is strict. One can easily check that S remains left-basic, overlap-free and finite. Moreover

$$L_1=L_2 \text{ iff } [\psi(f_1)]_{\xleftrightarrow{*}_S}=\psi[\varphi_1^{-1}(\#L_2\$)\cap R_1].$$

The left-hand side of this last equality is a class specified by a left-basic, overlap-free, strict, finite semi-Thue system while the right-hand side is a dcfl. □

Let us consider the following generalisation of the notion of syntactic congruence of a language. Given an alphabet X and two languages L_1, L_2 over X, we define the relation $\mathrm{Synt}_X(L_1, L_2)$ as: for every $f, g \in X^* \times X^*$, $(f, g) \in \mathrm{Synt}_X(L_1, L_2)$ iff $\forall \alpha \in L_1, \forall \beta \in X^*, \alpha f \beta \in L_2 \Leftrightarrow \alpha g \beta \in L_2$. This relation is a right-congruence. One can notice that $\mathrm{Synt}_X(X^*, L_2)$ is nothing other than the syntactic congruence of L_2.

Let us make one more step toward the reduction of Problem 1 to Problem 3.

5.19. Proposition. *The equivalence problem for one* dpda *and one class specified by a left-basic, overlap-free, strict, finite, proper semi-Thue system, reduces to the word problem for a right-congruence* $\mathrm{Synt}_X(R, [\sigma]_{\leftrightarrow_S})$ *defined by a rational, proper language* R*, a left-basic, overlap-free, strict, finite, proper semi-Thue system* S *and a letter* σ.

Proof. Let us consider a finite alphabet X, a dpda A_1 over X, a word $f_2 \in X^*$ and a left-basic, overlap-free, strict, finite semi-Thue system S_2 over X. We note $L_1 = L(A_1)$, $L_2 = [f_2]_{\leftrightarrow_{S_2}}$. By Theorem 3.1, $L_1 = [R]_{\leftrightarrow_{S_1}}$ where R is a rational subset of X^* and S_1 is a basic, confluent, strict, finite c-system.

$$S_1 = \bigcup_{i=1}^{n} R_i \times \{u_i\} \times \{v_i\}.$$

Let a, b, σ be new letters. We define

$$Y = X \cup \{a, b, \sigma\}, \qquad K = aRb \cup \{\sigma\}, \qquad K_i = aR_i$$

$$T_1 = \bigcup_{i=1}^{n} K_i \times \{u_i\} \times \{v_i\}, \qquad L_1' = aL_1 b \cup \{\sigma\}$$

$$T_2 = S_2 \cup \{(\sigma, af_2 b)\}, \qquad L_2' = aL_2 b \cup \{\sigma\}.$$

We have $L_1' = [K]_{\leftrightarrow_{T_1}}$ and $L_2' = [\sigma]_{\leftrightarrow_{T_2}}$ where T_1 is still a basic, confluent, strict, finite c-system, T_2 is a left-basic, overlap-free, strict, finite, proper semi-Thue system, K is rational and σ is a letter. Clearly, $L_1 = L_2$ iff $L_1' = L_2'$. We claim that this last equality is true iff

(i) $K \subset L_2'$,

(ii) $\mathrm{Irr}(T_1) - K \subset \bar{L}_2'$,

(iii) $\forall i \in [1, n]$, $(u_i, v_i) \in \mathrm{Synt}_Y(K_i, L_2')$.

Here $\bar{L}$ denotes the complement of L in X^*, for any language L over X.

Let us prove this claim. As T_1 is confluent, $\bar{L}_1' = [\mathrm{Irr}(T_1) - K]_{\leftrightarrow_{T_1}}$. We then always have $K \subset L_1'$, $\mathrm{Irr}(T_1) - K \subset \bar{L}_1'$, Moreover $\leftrightarrow_{T_1}$ saturates L_1', which is equivalent to

$$\forall i \in [1, n], \quad (u_i, v_i) \in \mathrm{Synt}_Y(K_i, L_1').$$

- If $L_1' = L_2'$, replacing L_1' by L_2' in the above affirmations we obtain properties (i), (ii), (iii).

- Let us suppose that properties (i), (ii), (iii) are true. By (iii) $\leftrightarrow_{T_1}$ saturates L'_2, hence $\overset{*}{\leftrightarrow}_{T_1}$ saturates L'_2. By (i), $K \subset L'_2$. As $\overset{*}{\leftrightarrow}_{T_1}$ saturates L'_2, it follows that $[K]_{\overset{*}{\leftrightarrow}_{T_1}} \subset L'_2$. By (ii), we obtain $[\mathrm{Irr}(T_1) - K]_{\overset{*}{\leftrightarrow}_{T_1}} \subset \bar{L}'_2$.
  Hence $L'_1 \subset L'_2$ and $\bar{L}'_1 \subset \bar{L}'_2$, which implies that $L'_1 = L'_2$. The claim is proved.

Conditions (i) and (ii) are decidable because K and $\mathrm{Irr}(T_1) - K$ are rational while L'_2 and $\bar{L}'_2$ are deterministic context-free. The problem $L(A_1) = [f_2]_{\overset{*}{\leftrightarrow}_{S_2}}$ is then reduced to a finite number of instances of the problem $(u, v) \in \mathrm{Synt}_Y(H, L)$ where H is a proper, rational subset of Y^* and L has the form $[\sigma]_{\overset{*}{\leftrightarrow}_S}$ for some letter $\sigma \in Y$ and some left-basic, overlap-free, strict, finite, proper semi-Thue system over Y. $\square$

The next step of our reduction of Problem 1 to Problem 3 is given by the following.

5.20. Proposition. *The word problem for a right-congruence* $\mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$ *defined by a rational, proper language R, a left-basic, overlap-free, strict, finite, proper, semi-Thue system S and a letter σ reduces to the word problem for the syntactic congruence of one class specified by a left-basic, confluent, strict, finite semi-Thue system.*

The proof of this proposition is similar to the proof of Proposition 5.1. It consists of defining a new alphabet Y which encodes the calculi of a finite automaton A which recognises both languages R and $\mathrm{Irr}(S)$. We then define a semi-Thue system T over Y which is made of the rules of S augmented by new rules encoding the reductions of S that overlap the words w_1, w_2 in a left-context belonging to R.

Let us fix some finite alphabet X, a rational proper language R over X, a letter $\sigma \in X$, a left-basic, overlap-free, strict, finite proper semi-Thue system S over X and two words $w_1, w_2 \in X^*$. As $\overset{*}{\leftrightarrow}_S \subset \mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$, $(w_1, w_2) \in \mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$ iff $(\theta_S(w_1), \theta_S(w_2)) \in \mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$. Hence we can always suppose that w_1, w_2 are S-irreducible.

We notice that if R is proper and S is proper, $\theta_S(R)$ is also a proper language. By Theorem 3.8, $\theta_S(R)$ is a rational language. One can check that the relations $\mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$ and $\mathrm{Synt}_X(\theta_S(R), [\sigma]_{\overset{*}{\leftrightarrow}_S})$ are equal. Hence we can always suppose that R is a subset of $\mathrm{Irr}(S)$.

Let $A = \langle X, Q, q_0, d, Q\rangle$ be a finite, complete, deterministic automaton which recognises R and $\mathrm{Irr}(S)$ in the sense that there are two subsets Q_+, Q_0 of Q such that $L(A, Q_+) = R$ and $L(A, Q_0) = \mathrm{Irr}(S)$. By the assumption that $R \subset \mathrm{Irr}(S)$, $Q_+ \subset Q_0$. We suppose that every state of A is accessible. We denote by Q_c the set of states which are Q_+-coaccessible:

$$Q_c = \{q \in Q, \exists f \in X^*, q.f \in Q_+\}.$$

We denote by Left(S) the set $\{u \in X^*, \exists v \in X^*, (u, v) \in S\}$. We define two integers

$$l = \max\{|u|\}_{u \in \mathrm{Left}(S)} \quad \text{and} \quad l' = \max\{|w_1|, |w_2|\}.$$

We say that a word u is a *left-atom for S* iff, for every $r, r', s' \in X^*$ and $(u', v') \in S$, $(ru = r'v's'$ and $|r| < |r'v'|) \Rightarrow s' = \varepsilon$. As S is overlap-free here, a word u is a left-atom for S iff for every $r \in \mathrm{Irr}(S)$, (r, u) is a left-block for S (see Definition 4.11). We denote by M the set of words

$$M = \{w \in X^* \mid w \text{ is } S\text{-irreducible and } w = uu' \text{ for some left-atom } u \text{ and some word } u' \text{ such that } |u'| \leq l' \text{ and } |u| + |u'| \leq l + l'\}.$$

5.21. Lemma. *Let $u, v', v'', s \in X^*$ such that $(u, v'v'') \in S$ and $v''s \in M$. Then $\theta_S(us) \in M$.*

Proof. Let us define

$$M' = \{w \in X^* \mid w = u'u'' \text{ where } u' \text{ is a left-atom and } u'' \text{ is a word such that } |u''| \leq l' \text{ and } |u'| + |u''| \leq l + l'\}.$$

One can check that $\vdash_S$ saturates M'. But $(u, v'v'') \in S$ and $v''s \in M$ imply that $us \in M'$. Hence $\theta_S(us) \in M' \cap \mathrm{Irr}(S) = M$. $\square$

Alphabet Y: The alphabet Y is made of the following symbols:

(1) all the letters $x \in X$,

(2) two special letters W_1, W_2,

(3) symbols of the form $[p, x, q]$, for every $p, q \in Q_c$, $x \in X$ such that $p.x = q$,

(4) symbols of the form $\overline{[p, w, q]}$, for every $p \in Q_c$, $q \in Q$, $w \in M$ such that $p.w = q$.

Let us define $\varphi : Y^* \to X^*$ as the homomorphism preserving X, sending W_k on w_k $(k \in \{1, 2\})$ and erasing the states. More precisely, if $x \in X$, $\varphi(x) = x$, if $k \in \{1, 2\}$, $\varphi(W_k) = w_k$, $\varphi([p, x, q]) = x$ and $\varphi(\overline{[p, w, q]}) = w$. Given a letter y of the form $[p, x, q]$ (resp. $\overline{[p, w, q]}$) we call p the *left-state of* y, and we call q the *right-state of* y. A word $f \in Y^*$ is said to be *monotonous* iff every factor of length 2, $y_1 y_2$, of f, is such that either $y_1 \in X \cup \{W_1, W_2\}$ or $y_2 \in X \cup \{W_1, W_2\}$ or the right-state of y_1 is equal to the left-state of y_2. Given a word $m \in X^*$ and two states $p, q \in Q$ such that $p.m = q$, we define $p \stackrel{m}{—} q$ as the unique word $f \in Y^*$ such that every letter of f is of type 3, f is monotonous, the left-state of the first letter of f is p and the right-state of the last letter of f is q.

System T: The semi-Thue system T is constituted of all the rules of the following types:

(1) $[p, x, q]\overline{[q, w_k, r]} \to [p, x, q]W_k$ where $q \in Q_+$, $k \in \{1, 2\}$,

(2) $\overline{[p, \theta_S(us), r']} \to p \stackrel{v'}{—} q\overline{[q, v''s, r]}$ where $(u, v'v'') \in S$, $v''s \in M$.

(3) $\overline{[p, s, q']}u \to \overline{[p, sv', q]}v''$ where $(u, v'v'') \in S$, $q \in Q_c$, $v' \neq \varepsilon$, $v'' \neq \varepsilon$, $sv' \in M$.

(4) $\overline{[p, \varepsilon, p]}u \to p \stackrel{v'}{—} q\overline{[q, v'', r]}v'''$ where $(u, v'v''v''') \in S$, $p \in Q_c$, $v' \neq \varepsilon$, $v''' \neq \varepsilon$, $v'' \in M$.

(5) $u \to v$ where $(u, v) \in S$.

Some precisions about T: By Lemma 5.21 if $(u, v'v'') \in S$ and $v''s \in M$, then $\theta_S(us) \in M$. Hence, every time when $p \stackrel{v'}{—} q$ and $[q, v''s, r]$ are defined and $(u, v'v'')$ is a rule of S, $\overline{[p, \theta_S(us), r']}$ is also defined, so that they lead to a rule of type 2.

5.22. Lemma. *T is overlap-free.*

Proof. For every $(i,j)\in[1,5]\times[1,5]$ we check that no pair of rules $((\alpha,\beta),(\alpha',\beta'))$, where (α,β) is of type i and (α',β') of type j, is overlapping. We omit the trivial cases.

$(1,j)$ (for any $j\in[1,5]$): No overlap of this kind is possible because the only right-hand sides of rules containing a letter W_k $(k\in\{1,2\})$ are of type 1 and such a right-hand side completely defines its left-hand side.

$(2,j)$: $(2,2)$ Such an overlap is impossible because S is overlap-free.

$(2,3)$ Here $\beta=p\xrightarrow{v'}q[q,v''s,r]$, $\beta'=[p',s'v''',q']v''''$. The fact that $q'\in Q_0$ while $r\notin Q_0$ (because p is accessible and $v'v''$ is a redex of S) implies that (β,β') is not overlapping.

$(2,4)$ Such an overlap is impossible because S is overlap-free.

$(3,j)$: $(3,2)$ Same argument as for case $(2,4)$.

$(4,j)$ or $(5,j)$: Here also, every overlap would lead to an overlap in S, which is overlap-free. $\square$

5.23. Lemma. *T is left-basic.*

Proof. The same kind of arguments as in the proof of Lemma 5.21 can be used to treat all the cases except the following: (α,β) is of type (3) while (α',β') is of type (2). Let us show that α cannot be overlapped on the left by β'.

$$\alpha=\overline{[p,s,q']}u,\qquad \beta=\overline{[p,sv',q]}v''\quad\text{where } q\in Q_0,$$

$$\beta'=p'\xrightarrow{v'''}q''\overline{[q'',v''''s',r]}\qquad\qquad\text{where } (u',v'''v'''')\in S,$$

$(p.s).v'=q\in Q_0$ and p is accessible, hence $p.s\in Q_0$. So, $q'\in Q_0$ while $r\notin Q_0$. It follows that the letters $[p,s,q']$ and $[q'',v''''s',r]$ are different, so that α and β' cannot overlap. $\square$

Let us consider the following valuation $\nu: Y^*\to\mathbb{N}$ defined by

$$\forall x\in X,\ \nu(x)=1;\qquad \forall k\in\{1,2\},\ \nu(W_k)=l'+2$$

$$\text{for every letter } [p,x,q],\quad \nu([p,x,q])=1$$

$$\text{for every letter } \overline{[p,w,q]},\quad \nu(\overline{[p,w,q]})=|w|+1.$$

5.24. Lemma. *T is* v-*strict with respect to ν.*

The proof is trivial.

5.25. Lemma. *Let us consider the finite set*

$$F=\{[q_0,\sigma,q_1]\overline{[q_1,\varepsilon,q_1]},\ \overline{[q_0,\sigma,q_1]},\ \overline{[q_0,\varepsilon,q_0]}\sigma\}\quad \textit{where } q_0.\sigma=q_1.$$

For every $h\in Y^*$, $h\in[F]_{\leftrightarrow_T}$ iff the three following conditions on h hold

(1) *h can be decomposed as $h=fWg$ where $f=q_0\xrightarrow{m}q$, $W\in\{W_1,W_2,\overline{[q,w,r]}\}$, $g\in X^*$ for some $m\in X^*$, $q\in Q_0$, $w\in M$, $r\in Q$,*

(2) *if $W\in\{W_1,W_2\}$, then $q\in Q_+$,*

(3) *$\varphi(h)\in[\sigma]_{\leftrightarrow_S}$.*

Proof. Let $L = \{h \in Y^*$, conditions (1), (2), (3) on h hold$\}$. We shall prove that

(a) if $h \in L - F$, then h is T-reducible,

(b) $\vdash_T$ saturates F,

(c) $\to_T$ saturates L.

From these properties and the fact that $F \subset L$, Lemma 5.25 will follow.

(a) Let $h \in L - F$. We consider a decomposition $h = fWg$ satisfying conditions (1), (2), (3).

Case 1: $W \in \{W_1, W_2\}$. By condition (2), $q \in Q_+$. As R is proper, we must have $m \neq \varepsilon$ and hence $f \neq \varepsilon$. The last letter of f is of the type $[q', x, q]$ and hence h is T-reducible (by a rule of type 1).

Case 2: $W \notin \{W_1, W_2\}$ and $\varphi(h)$ is S-irreducible. By condition (3) and the hypothesis that $\varphi(h)$ is S-irreducible, $\varphi(h) = \sigma$. It follows that $h \in F$, contradicting the assumption that $h \in L - F$. Hence this case is impossible.

Case 3: $W \notin \{W_1, W_2\}$ and $\varphi(h)$ is S-reducible. So, $h = f\overline{[q, w, r]}g$ where $f = q_0 \overset{m}{\longrightarrow} q$, $m \in X^*$, $q \in Q_c$, $w \in M$, $r \in Q$, $g \in X^*$.

$\varphi(h) = mwg = \alpha v \beta$ for some $\alpha, \beta \in X^*$ and $(u, v) \in S$ such that this occurrence of v is a leftmost redex. As $q \in Q_c$, m is S-irreducible and by definition of M, w is S-irreducible. Four subcases can arise depending on the position of the occurrence of v.

Subcase 1 (see Fig. 10(a)): Here $m = \alpha v'$, $w = v''s$ where $v'v'' = v$, $\alpha, s \in X^*$. Hence h is T-reducible by a rule of type (2).

Subcase 2 (see Fig. 10(b)): Here $w = sv'$, $g = v''g'$ where $v'v'' = v$, $v'' \neq \varepsilon$, $s, g' \in X^*$. If $v' = \varepsilon$, then h is T-reducible by a rule of type (5). If $v' \neq \varepsilon$, then, as the occurrence of v is a leftmost redex, $q.sv' \in Q_0$. So, h is T-reducible by a rule of type (3).

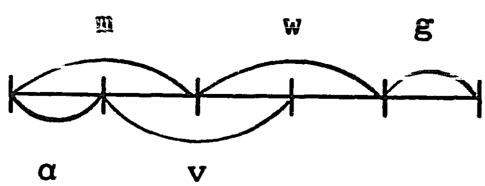

Fig. 10(a).

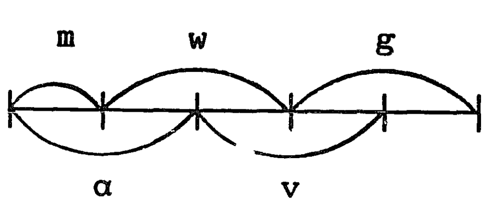

Fig. 10(b).

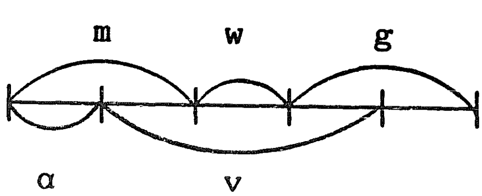

Fig. 10(c).

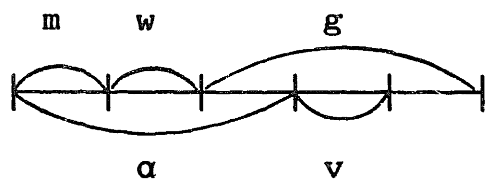

Fig. 10(d).

Subcase 3 (see Fig. 10(c)): $m = \alpha v'$, $w = v''$, $g = v'''s$ where $v'v''v''' = v$, $s \in X^*$, $v' \neq \varepsilon$, $v''' \neq \varepsilon$. Here h is T-reducible by a rule of type (4).

Subcase 4 (see Fig. 10(d)): $g = g'vg''$ where $g', g'' \in X^*$. Here h is T-reducible by a rule of type (5).

Point (a) is then proved.

(b) The conjunction of conditions (1) and (2) is preserved by $\vdash_T$. For every rule $(\alpha, \beta) \in T$, $\varphi(\beta) \overset{*}{\vdash}_T \varphi(\alpha)$. Hence condition (3) is preserved by $\vdash_T$.

(c) By the same means one can prove that L is saturated by $\rightarrow_T$. $\square$

5.26. Lemma. $(w_1, w_2) \in \mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$ *iff* $(W_1, W_2) \in \mathrm{Synt}_Y(Y^*, [F]_{\overset{*}{\leftrightarrow}_T})$.

Proof. Let us denote by $\mathscr{C}_k$ $(k \in \{1, 2\})$ the set of contexts in X^* sending w_k in $[\sigma]_{\overset{*}{\leftrightarrow}_S}$ (i.e. the set of pairs $(\alpha, \beta) \in X^* \times X^*$ such that $\alpha w_k \beta \in [\sigma]_{\overset{*}{\leftrightarrow}_S}$). Similarly $\bar{\mathscr{C}}_k$ $(k \in \{1, 2\})$ denotes the set of contexts in Y^* sending W_k in $[F]_{\overset{*}{\leftrightarrow}_T}$. Let

$K = \{f \in Y^* \mid$ every letter of f is of type (3) and the first letter of f has a left state equal to q_0 and the last letter of f has a right state which belongs to $Q_+\}$.

Let $\bar{\varphi} : Y^* \times Y^* \to X^* \times X^*$ be the homomorphism defined by

$$\bar{\varphi}((\alpha, \beta)) = (\varphi(\alpha), \varphi(\beta)).$$

Lemma 5.25 shows that for every $k \in \{1, 2\}$

$$\bar{\mathscr{C}}_k = \bar{\varphi}^{-1}(\mathscr{C}_k) \cap K \times X^*, \qquad \mathscr{C}_k = \bar{\varphi}(\bar{\mathscr{C}}_k).$$

Hence $\mathscr{C}_1 = \mathscr{C}_2$ iff $\bar{\mathscr{C}}_1 = \bar{\mathscr{C}}_2$. $\square$

Proof of Proposition 5.20. (*Conclusion*). Let Y' be a new alphabet defined by $Y' = Y \cup \{a, b, \sigma'\}$ where a, b, σ' are new letters. Let $T' = T \cup \{(\sigma', afb)\}_{f \in F}$. From lemmas 5.22 and 5.25 and the fact that a, b are new letters we conclude that T' is overlap-free and left-basic. As $[\sigma']_{\overset{*}{\leftrightarrow}_{T'}} = a([F]_{\overset{*}{\leftrightarrow}_T})b \cup \{\sigma'\}$, using Lemma 5.26, we see that

$$(w_1, w_2) \in \mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S}) \textit{ iff } (W_1, W_2) \in \mathrm{Synt}_{Y'}(Y'^*, [\sigma]_{\overset{*}{\leftrightarrow}_{T'}}).$$

If we extend the valuation ν, used in lemma 5.24, to the alphabet Y' by setting $\nu(a) = \nu(b) = \nu(\sigma') = 1$, then T' is v-strict with respect to ν. Using the same trick as in the proof of Proposition 5.18 one can define a new alphabet $\hat{Y}'$ and a homomorphism $\psi: Y'^* \to \hat{Y}'^*$ such that $\psi(T')$ is still left-basic, overlap-free, finite and, in addition $\psi(T')$ is strict. We also have $(W_1, W_2) \in \mathrm{Synt}_{Y'}(Y'^*, [\sigma']_{\overset{*}{\leftrightarrow}_{T'}})$ iff

$$(\psi(W_1), \psi(W_2)) \in \mathrm{Synt}_{\hat{Y}'}(\hat{Y}'^*, [\sigma']_{\overset{*}{\leftrightarrow}_{\psi(T')}})$$

(see the proof of Proposition 5.18).

Hence we are reduced to the word-problem for the syntactic congruence of one class specified by a left-basic, confluent, strict, finite semi-Thue system. □

Proof of Theorem 5.17. Let us denote by $P \leqslant P'$ the fact that the problem P recursively reduces to the problem P'.

Problem $1 \leqslant$ *Problem* 3: This is a consequence of Propositions 5.18, 5.19, 5.20.

Problem $1 \leqslant$ *Problem* 2: By Propositions 5.18, 5.19, problem P1 reduces to the word problem for a right-congruence $\mathrm{Synt}_X(R, [\sigma]_{\overset{*}{\leftrightarrow}_S})$ defined by a rational, proper language R, a left-basic, overlap-free, strict, finite, proper semi-Thue system S and a letter σ. Let us start now with this last problem and use the notations and the constructions given in the proof of Proposition 5.20. Let us look at the semi-Thue system T. Let $\eta: Y^* \to Y^*$ be the homomorphism which preserves every letter of $Y - \{W_1, W_2\}$ and which exchanges the letters W_1, W_2:

$$\forall y \in Y - \{W_1, W_2\}, \eta(y) = y, \qquad \eta(W_1) = W_2, \qquad \eta(W_2) = W_1.$$

Let us denote by $\eta(T)$ the semi-Thue system $\{(\eta(u), \eta(v))\}_{(u,v) \in T}$. We claim that $(W_1, W_2) \in \mathrm{Synt}_Y(Y^*, [F]_{\overset{*}{\leftrightarrow}_T})$ if and only if $[F]_{\overset{*}{\leftrightarrow}_T} = [F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$.

(1) Let us suppose that $(W_1, W_2) \in \mathrm{Synt}_Y(Y^*, [F]_{\overset{*}{\leftrightarrow}_T})$. Let us denote by $\equiv$ the syntactic congruence of $[F]_{\overset{*}{\leftrightarrow}_T}$. As $W_1 \equiv W_2$, for every $w \in Y^*$, $w \equiv \eta(w)$. Hence $\eta([F]_{\overset{*}{\leftrightarrow}_T}) \subset [F]_{\overset{*}{\leftrightarrow}_T}$ and, as $\eta \circ \eta = \mathrm{Id}_{Y^*}$, $[F]_{\overset{*}{\leftrightarrow}_T} \subset \eta([F]_{\overset{*}{\leftrightarrow}_T})$. As $\eta(F) = F$ we have also the equality $\eta([F]_{\overset{*}{\leftrightarrow}_T}) = [F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$. Finally we have proved that $[F]_{\overset{*}{\leftrightarrow}_T} = [F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$.

(2) Let us suppose that $[F]_{\overset{*}{\leftrightarrow}_T} = [F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$. Let $k \in \{1, 2\}$ and $\alpha, \beta \in Y^*$ such that $\alpha W_k \beta \in [F]_{\overset{*}{\leftrightarrow}_T}$. Then $\eta(\alpha)\eta(W_k)\eta(\beta) \in [F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$. By Lemma 5.25, α, β are in $(Y - \{W_1, W_2\})^*$, hence $\eta(\alpha) = \alpha$, $\eta(\beta) = \beta$. So, $\alpha\eta(W_k)\beta$ is in $[F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$ which by hypothesis, is the language $[F]_{\overset{*}{\leftrightarrow}_T}$. Hence $W_k \equiv \eta(W_k)$.

Our claim is proved. By this claim and Lemma 5.26, Problem 1 is now reduced to the problem $[F]_{\overset{*}{\leftrightarrow}_T} = [F]_{\overset{*}{\leftrightarrow}_{\eta(T)}}$. One can notice that the sets $\{u \in Y^*, \exists v \in Y^*, (u, v) \in T\} = \mathrm{Left}(T)$ and $\{y \in Y^*, \exists u \in Y^*, (u, v) \in T\} = \mathrm{Right}(T)$ are preserved by η. Hence T and $\eta(T)$ are both left-basic and overlap-free. η preserves ν, hence $\eta(T)$ is v-strict with respect to ν. Let us extend η to Y' by setting $\forall y \in \{a, b, \sigma\}$, $\eta(y) = y$. It is straightforward that

$$[F]_{\overset{*}{\underset{T}{\leftrightarrow}}} = [F]_{\overset{*}{\underset{\eta(T)}{\leftrightarrow}}} \text{ iff } [\sigma']_{\overset{*}{\underset{\psi(T')}{\leftrightarrow}}} = [\sigma']_{\overset{*}{\underset{\psi\circ\eta(T')}{\leftrightarrow}}}.$$

Testing this last equality is an instance of Problem 2.

Problem $1 \leqslant$ *Problem* 4: It suffices to reduce the instance of Problem 2 obtained in the reduction above to an instance of Problem 4. Let $T'' = \psi(T') \cup \psi \circ \eta(T')$. We claim that $[\sigma']_{\overset{*}{\leftrightarrow}_{\psi(T')}} = [\sigma']_{\overset{*}{\leftrightarrow}_{\psi\circ\eta(T')}}$ iff $[\sigma']_{\overset{*}{\leftrightarrow}_{T''}} = \langle\sigma'\rangle_{\overset{*}{\leftrightarrow}_{T''}}$.

(1) Let us suppose that $[\sigma']_{\overset{*}{\leftrightarrow}_{\psi(T')}} = [\sigma']_{\overset{*}{\leftrightarrow}_{\psi\circ\eta(T')}}$. Let us note $L = [\sigma']_{\overset{*}{\leftrightarrow}_{\psi(T')}}$. Both relations $\leftrightarrow_{\psi(T')}$, $\leftrightarrow_{\psi\circ\eta(T')}$ are saturating L. Hence their union, $\leftrightarrow_{T''}$ is saturating L. Hence $[\sigma']_{\overset{*}{\leftrightarrow}_{T''}} \subset L$. As $\psi(T')$ is confluent $L = \langle\sigma'\rangle_{\overset{*}{\rightarrow}_{\psi(T')}}$, which is included in $\langle\sigma'\rangle_{\overset{*}{\rightarrow}_{T''}}$. Hence $[\sigma']_{\overset{*}{\leftrightarrow}_{T''}} \subset \langle\sigma'\rangle_{\overset{*}{\rightarrow}_{T''}}$. This implies that $[\sigma']_{\overset{*}{\leftrightarrow}_{T''}} = \langle\sigma'\rangle_{\overset{*}{\rightarrow}_{T''}}$.

(2) Let us suppose that $[\sigma']_{\overset{*}{\leftrightarrow}_{T''}} = \langle\sigma'\rangle_{\overset{*}{\rightarrow}_{T''}}$. Let $f \in [\sigma']_{\overset{*}{\leftrightarrow}_{\psi(T')}}$. Let $\bar{f}$ be the unique word such that $f \vdash^{*}_{\psi\circ\eta(T')} \bar{f}$ and $\bar{f}$ is $\psi \circ \eta(T')$-irreducible. $\bar{f} \in [\sigma']_{\overset{*}{\leftrightarrow}_{T''}}$ and, as: Right$(\eta(T))$ = Right(T), we also have Right$(\psi \circ \eta(T'))$ = Right$(\psi(T'))$. Hence $\bar{f}$ is T''-irreducible. By the hypothesis, $\bar{f} \in \langle\sigma'\rangle_{\overset{*}{\rightarrow}_{T''}}$, so $\bar{f} = \sigma'$ and $f \in [\sigma']_{\overset{*}{\leftrightarrow}_{\psi\circ\eta(T')}}$. We have proved the inclusion $[\sigma']_{\overset{*}{\leftrightarrow}_{\psi(T')}} \subset [\sigma']_{\overset{*}{\leftrightarrow}_{\psi\circ\eta(T')}}$. The converse inclusion can be proved by same means. Our claim is then proved.

As Left(T'') = Left$(\psi(T'))$ and Right(T'') = Right$(\psi(T'))$, T'' is left-basic and strict (but it contains overlaps of type II). Testing whether $[\sigma']_{\overset{*}{\leftrightarrow}_{T''}} = \langle\sigma'\rangle_{\overset{*}{\rightarrow}_{T''}}$ is hence an instance of Problem 4.

Problem $2 \leqslant$ *Problem* 1: This is clear because every language of the form $[f]_{\overset{*}{\leftrightarrow}_S}$ where S is a left-basic, confluent, strict and finite semi-Thue system is recognised by some dpda which is computable from f and S.

Problem $3 \leqslant$ *Problem* 1: Let us consider a finite alphabet X, three words f, g, $h \in X^*$ and one semi-Thue system S over X, such that S is left-basic, confluent, strict and finite. Let us note $L = [h]_{\overset{*}{\leftrightarrow}_S}$. Problem 3 is "$f \equiv_L g$?" We consider a new letter # and for every word $w \in X^*$ we set

$$L_w = \{\alpha\#\beta \mid \alpha \in X^*, \beta \in X^*, \alpha w \beta \in L\}.$$

$f \equiv_L g$ iff $L_f = L_g$, and L_f, L_g are dcfls. Hence we are reduced to an instance of Problem 1.

Problem $4 \leqslant$ *Problem* 1: Let us consider a finite alphabet X, a word $f \in X^*$ and a semi-Thue system S over X such that S is left-basic, strict and finite. Problem 4 is "$\langle f\rangle_{\overset{*}{\rightarrow}_S} = [f]_{\overset{*}{\leftrightarrow}_S}$?" Let us denote by $\overset{*}{\rightarrow}_S$ the inverse of relation $\vdash^{*}_{\shortmid S}$: $g \overset{*}{\rightarrow}_{\shortmid S} h$ iff $h \vdash^{*}_{\shortmid S} g$. Let $L = \langle f\rangle_{\overset{*}{\rightarrow}_{\shortmid S}}$. We claim that $\langle f\rangle_{\overset{*}{\rightarrow}_S} = [f]_{\overset{*}{\leftrightarrow}_S}$ iff, for every $(u, v) \in S$, $u \equiv_L v$.

(1) Let us suppose that $\langle f\rangle_{\overset{*}{\rightarrow}_S} = [f]_{\overset{*}{\leftrightarrow}_S}$. Let us show that $\leftrightarrow_S$ saturates $\langle f\rangle_{\overset{*}{\rightarrow}_{\shortmid S}}$. Let $g \in \langle f\rangle_{\overset{*}{\rightarrow}_{\shortmid S}}$ and $h \in X^*$ such that $g \leftrightarrow_S h$. There exists some word $\bar{h}$ such that $h \vdash^{*}_{\shortmid S} \bar{h}$ and $\bar{h}$ is S-irreducible. $f \overset{*}{\leftrightarrow}_S g \leftrightarrow_S h \overset{*}{\leftrightarrow}_S \bar{h}$ hence $\bar{h} \in [f]_{\overset{*}{\leftrightarrow}_S}$ which is equal to $\langle f\rangle_{\overset{*}{\rightarrow}_S}$. As $\bar{h}$ is S-irreducible, $\bar{h} = f$. Hence $h \in \langle f\rangle_{\overset{*}{\rightarrow}_{\shortmid S}}$. It is clear that $\leftrightarrow_S$ saturates $\langle f\rangle_{\overset{*}{\rightarrow}_{\shortmid S}}$ iff, for every $(u, v) \in S$, $u \equiv_L v$.

(2) Let us suppose that for every $(u, v) \in S$, $u \equiv_L v$. By this hypothesis, $\leftrightarrow_S$ saturates L. $f \in L$ and $\leftrightarrow_S$ saturates L, so $[f]_{\overset{*}{\leftrightarrow}_S} \subset L$. $L = \langle f\rangle_{\overset{*}{\rightarrow}_{\shortmid S}} \subset \langle f\rangle_{\overset{*}{\rightarrow}_S}$. Hence $[f]_{\overset{*}{\leftrightarrow}_S} \subset \langle f\rangle_{\overset{*}{\rightarrow}_S}$ and, as the converse inclusion is trivial, $\langle f\rangle_{\overset{*}{\rightarrow}_S} = [f]_{\overset{*}{\leftrightarrow}_S}$. The claim is proved.

The relation $\vdash_{\shortmid S}$ is generated by the left-basic, strongly injective, strict, finite c-system LMR(S). Hence, by Theorem 3.1, L is a dcfl. Now, by the trick used in the reduction of Problem 3 to Problem 1, every question "$u \equiv_L v$?" reduces to a question of the form "$L_u = L_v$?" where L_u, L_v are dcfls. Hence every instance of Problem 4 reduces to a finite number of instances of Problem 1. □

Problem 2 and Problem 3 have already been shown to be decidable for basic (instead of left-basic), confluent, strict, finite semi-Thue systems [34]. Here we show that Problem 4 is also decidable for basic (instead of left-basic) strict, finite semi-Thue systems. We prove the following more general result, which deals with partial confluence on a rational set instead of partial confluence on a single word.

5.27. Theorem. *The problem of partial confluence on a rational set is decidable for basic, strict, finite semi-Thue systems.*

Let us define more precisely this problem (we shall call it Problem 4′).

Instance: One rational set R and one semi-Thue system S over a finite alphabet X, such that S is basic, strict and finite.

Question: $\langle R\rangle_{\to_S} = [R]_{\leftrightarrow_S}$?

Let us consider an instance (R, S, X) of Problem 4′. Let # be a new letter. For every $w \in X^*$ we define

$$\mathrm{IC}(w) = \{\alpha \# \beta \mid \alpha \in \mathrm{Irr}(S), \beta \in \mathrm{Irr}(S), \alpha w \beta \in \langle R\rangle_{\vdash^{-1}_S}\}.$$

The letters IC stand for irreducible contexts. The reader will notice that the main difference between IC(w) and the language L_w defined in the reduction of Problem 3 to Problem 1 is that IC(w) encodes only the *irreducible contexts sending* w in $\langle R\rangle_{\to_S}$.

We recall that the equivalence problem for finite-turn dpda is solved by the so-called Parallel Stacking Algorithm (this algorithm is defined in [38] and its complexity is analysed in [2]). Let us denote by PSA(A_1, A_2) the result of this algorithm applied on the finite-turn dpda A_1, A_2 (this result is 1 if A_1, A_2 are equivalent, 0 otherwise).

Let us consider the following.

Algorithm

(1) RES := 1

(2) FOR every $(u, v) \in S$, DO

(3) Compute two one-turn dpda A_u, A_v, such that:

$$L(A_u) = \mathrm{IC}(u) \text{ and } L(A_v) = \mathrm{IC}(v)$$

(Such a computation is described in the proof of Proposition III-2 of [34]. Though this reference deals with a semi-Thue system P which is the set of productions of a proper cf grammar, it is still valid for a basic, strict, semi-Thue system S.)

(4) IF PSA(A_u, A_v) = 0 THEN RES := 0

(5) ROF

(6) WRITE (RES)

It is clear that this algorithm always terminates. Let us prove that it is correct in the sense that its output is 1 if and only if

$$\langle R\rangle_{\overset{*}{\underset{S}{\rightarrow}}}=[R]_{\overset{*}{\underset{S}{\leftrightarrow}}}$$

Proof. Our algorithm will be correct iff the following equivalence holds:

$$\langle R\rangle_{\overset{*}{\underset{S}{\rightarrow}}}=[R]_{\overset{*}{\underset{S}{\leftrightarrow}}} \Leftrightarrow \forall(u,v)\in S, \qquad \mathrm{IC}(u)=\mathrm{IC}(v).$$

(1) Let us suppose that $\langle R\rangle_{\overset{*}{\rightarrow}_S}=[R]_{\overset{*}{\leftrightarrow}_S}$. By the same means as in the reduction of Problem 4 to problem 1 (point (1)) one can then deduce that $\forall(u,v)\in S$, $u\equiv_L v$ where $L=\langle R\rangle_{\overset{*}{\rightarrow}|_S}$. This implies that for every $(u,v)\in S$, $\mathrm{IC}(u)=\mathrm{IC}(v)$.

(2) Let us suppose that for every $(u,v)\in S$, $\mathrm{IC}(u)=\mathrm{IC}(v)$. Let us denote by I the set $\mathrm{Irr}(S)$. We show that $L=\langle R\rangle_{\overset{*}{\rightarrow}|_S}$ and $\bar{L}=\langle I-R\rangle_{\overset{*}{\rightarrow}|_S}$ are both saturated by $\vdash_S$. If this was not true, then, there would exist a shortest word $h\in X^*$ such that there exists $h'\in X^*$ with $h\vdash_S h'$ and $(h,h')\in L\times\bar{L}\cup\bar{L}\times L$. In this word h, we consider some rightmost redex leading to a word h' fulfilling the above property: $h=\alpha v\beta$ where $(u,v)\in S$, $\alpha,\beta\in X^*$, $h'=\alpha u\beta$ is such that $(h,h')\in L\times\bar{L}\cup\bar{L}\times L$ and for every decomposition $h=\alpha'v'\beta'$, with $(u',v')\in S$ and $|\alpha|<|\alpha'|$, $(h,\alpha'u'\beta')\in L\times L\cup\bar{L}\times\bar{L}$.

As L, R and $\bar{L}$, $I-R$ play the same role we can assume, for example, that $(h,h')\in L\times\bar{L}$. We consider the unique word $\bar{\alpha}$ (resp. $\bar{\beta}$) which is S-irreducible and such that $\alpha\overset{*}{\vdash}_{|S}\bar{\alpha}$ (resp. $\beta\vdash_{|S}\bar{\beta}$).

5.28. Fact. *$\alpha v\bar{\beta}\overset{*}{\vdash}_{|S} r$ for some $r\in R$.*

As v is the rightmost redex leading from h to a word of $\bar{L}$, the leftmost redex occurring in β leads from h to a word $\alpha v\beta'\in L$. By minimality of h, as $\alpha v\beta'\in L$ any word h'' such that $\alpha v\beta'\overset{*}{\vdash}_S h''$ belongs to L. Hence $\alpha v\bar{\beta}\in L$, which proves Fact 5.28.

5.29. Fact. *$\bar{\alpha}v\bar{\beta}\overset{*}{\vdash}_{|S} r$.*

We know that $\alpha v\bar{\beta}\overset{*}{\vdash}_{|S} r$. As $\alpha\overset{*}{\vdash}_{|S}\bar{\alpha}$, $\alpha v\bar{\beta}\overset{*}{\vdash}_{|S}\bar{\alpha}v\bar{\beta}$. But the relation $\vdash_{|S}$ is generated by the strongly-injective c-system $\mathrm{LMR}(S)$. Hence the reduction $\alpha v\bar{\beta}\overset{*}{\vdash}_{|S}\bar{\alpha}v\bar{\beta}$ must be a prefix of the reduction $\alpha v\bar{\beta}\overset{*}{\vdash}_{|S} r$. So $\bar{\alpha}v\bar{\beta}\overset{*}{\vdash}_{|S} r$. Fact 5.29 is proved.

5.30. Fact. *$\bar{\alpha}u\bar{\beta}\overset{*}{\vdash}_{|S} r'$ for some $r'\in R$.*

By Fact 5.29, $\bar{\alpha}\#\bar{\beta}\in\mathrm{IC}(v)$ and $\mathrm{IC}(v)=\mathrm{IC}(u)$. Fact 5.30 follows.

Let us distinguish two cases.

Case 1: $\bar{a}u\beta \in L$. Then $\alpha u\beta \in L$, which is impossible because $\alpha u\beta = h' \in \bar{L}$.

Case 2: $\bar{a}u\beta \in \bar{L}$. $\bar{a}u\beta \vdash^{*}_{S} \bar{a}u\bar{\beta}$ and by minimality of h, every step of this reduction preserves $\bar{L}$. Hence $\bar{a}u\bar{\beta} \in \bar{L}$, contradicting Fact 5.30. Hence, we have proved that L and $\bar{L}$ are both saturated by $\vdash_S$. This implies that L is saturated by $\leftrightarrow_S$, so that $[R]_{\leftrightarrow_S} \subset L$. We have then the relations

$$[R]_{\overset{*}{\underset{S}{\leftrightarrow}}} \subset L = \langle R\rangle_{\overset{*}{\to}_{1_S}} \subset \langle R\rangle_{\overset{*}{\underset{S}{\to}}} \subset [R]_{\overset{*}{\underset{S}{\leftrightarrow}}}.$$

Hence $\langle R\rangle_{\overset{*}{\to}_S} = [R]_{\overset{*}{\leftrightarrow}_S}$. $\square$

5.31. Remark. (1) It is clear that our algorithm remains valid for basic, v-*strict*, finite semi-Thue systems.

(2) Let $G = \langle X, V, P, A\rangle$ be a context-free grammar (where X is the terminal alphabet, V the set of non-terminals, P the set of productions and A, the set of axioms, is included in V). G is proper iff for every rule $(v, w) \in P$, $w \neq \varepsilon$. A grammar G is said to be NTS [6, 8] iff for every $v \in V$, $\langle v\rangle_{\overset{*}{\to}_P} = [v]_{\overset{*}{\leftrightarrow}_P}$. As P is basic (because $\mathrm{Left}(P) \subset V$) and v-strict (if G is proper), the algorithm described here can be used to test whether G is NTS or not. Another algorithm testing whether a cf grammar is NTS or not is given in [34, Part IV, Proof of Proposition 3].

Let us prove now that Problem 4 becomes undecidable if we remove the hypothesis "left-basic".

5.32. Proposition. *The problem of partial confluence on a single word $\{f\}$ is undecidable for strict, finite semi-Thue systems.*

The next two lemmas will enable us to reduce the class equivalence problem for confluent, strict, finite semi-Thue systems to the partial confluence problem for strict, finite semi-Thue systems. As the first problem is undecidable (Proposition 4.7) it will follow that the second problem is also undecidable.

5.33. Lemma. *Let f be a word and S_1, S_2 be two semi-Thue systems on a finite alphabet X. $[f]_{\overset{*}{\leftrightarrow}_{S_1}} = [f]_{\overset{*}{\leftrightarrow}_{S_2}}$ iff*

(1) $\forall (u, v) \in S_1$, $(u, v) \in \mathrm{Synt}_X(X^*, [f]_{\overset{*}{\leftrightarrow}_{S_2}})$,

(2) $\forall (u, v) \in S_2$, $(u, v) \in \mathrm{Synt}_X(X^*, [f]_{\overset{*}{\leftrightarrow}_{S_1}})$.

The proof is left to the reader.

This lemma gives a reduction of the class equivalence problem for confluent, strict, finite semi-Thue systems to the word problem for the syntactic congruence of one class specified by a confluent, strict, finite semi-Thue system.

Let us consider an instance of this last problem, that is a finite alphabet X, three words $w_1, w_2, f \in X^*$ and one semi-Thue system S over X, such that S is confluent, strict, finite. The question is : $(w_1, w_2) \in \mathrm{Synt}_X(X^*, [f]_{\overset{*}{\leftrightarrow}_S})$? We can suppose that f is S-irreducible. Let W be a new letter. We define $Y = X \cup \{W\}$ and $\hat{S} = S \cup \{(w_1, W), (w_2, W)\}$. $\hat{S}$ is v-strict with respect to the valuation ν defined by

$$\forall x \in X, \quad \nu(x) = 1 \quad \text{and} \quad \nu(W) = \max\{|w_1|, |w_2|\} + 1.$$

5.34. Lemma. *Let* $L=[f]_{\leftrightarrow_S}$. $w_1 \equiv_L w_2$ *iff* $\langle f\rangle_{\leftarrow_S}=[f]_{\leftrightarrow_S}$.

Proof. (1) Let us suppose that $w_1 \equiv_L w_2$. To prove that $\hat{S}$ is partially confluent on $\{f\}$ it is enough to prove that $\langle f\rangle_{\leftarrow_{\hat{S}}}$ is saturated by $\vdash_{\hat{S}}$. Let g, $h \in Y^*$ such that $g \in \langle f\rangle_{\leftarrow_{\hat{S}}}$ and $g \vdash_{\hat{S}} h$. $g = u_0 W u_1 \ldots W u_n$ where the words u_i ($i \in [0, n]$) belong to X^*. There exists a word $\bar{g} = u_0 w_{j_1} u_1 \ldots w_{j_n} u_n$ (where for every $i \in [1, n]$, $j_i \in \{1, 2\}$) such that $g \models^*_{\hat{S}} \bar{g} \models^*_S f$.

Case 1: $g \vdash_S h$. Then $h = u_0 W \ldots u_{i-1} W u_i' W \ldots W u_n$ for some $i \in [0, n]$ where $u_i \vdash_S u_i'$. Let us define $\bar{h} = u_0 w_{j_1} \ldots u_{i-1} w_{j_i} u_i' w_{j_{i+1}} \ldots w_{j_n} u_n$. As $\bar{h} \leftrightarrow_S \bar{g}$ and S is confluent, $\bar{h} \models^*_S f$. Hence $h \models^*_{\hat{S}} \bar{h} \models^*_{\hat{S}} f$.

Case 2: $g \not\vdash_S h$. Hence there exists some $i \in [1, n]$ and some $k \in \{1, 2\}$ such that $h = u_0 W u_i \ldots W u_{i-1} w_k u_i W \ldots W u_n$. We define $\bar{h} = u_0 w_{j_1} u_1 \ldots w_{j_{i-1}} u_{i-1} w_k u_i w_{j_{i+1}} \ldots w_{j_n} u_n$. By hypothesis $w_{j_i} \equiv_L w_k$, hence $\bar{g} \equiv_L \bar{h}$ so $\bar{h} \in [f]_{\leftrightarrow_S}$ and by confluence of S, $\bar{h} \models^*_S f$. Hence $h \models^*_{\hat{S}} \bar{h} \models^*_{\hat{S}} f$.

(2) Let us suppose that $\langle f\rangle_{\leftarrow_{\hat{S}}}=[f]_{\leftrightarrow_{\hat{S}}}$. Let p, $s \in X^*$, $pw_1s \leftrightarrow_{\hat{S}} pWs \leftrightarrow_{\hat{S}} pw_2s$. Hence $pw_1s \in [f]_{\leftrightarrow_{\hat{S}}}$ iff $pw_2s \in [f]_{\leftrightarrow_{\hat{S}}}$. As $[f]_{\leftrightarrow_{\hat{S}}} = \langle f\rangle_{\leftarrow_{\hat{S}}}$ and $pw_ks \in X^*$ ($k \in \{1, 2\}$), $pw_ks \in [f]_{\leftrightarrow_{\hat{S}}}$ iff $pw_ks \in \langle f\rangle_{\leftarrow_S}$. Finally, $pw_1s \in \langle f\rangle_{\leftarrow_S}$ iff $pw_2s \in \langle f\rangle_{\leftarrow_S}$. Thus, $w_1 \equiv_L w_2$. □

Proof of Proposition 5.32. By Lemmas 5.33 and 5.34 we have reduced the class equivalence problem for confluent, strict, finite semi-Thue systems to the partial confluence problem for v-strict, finite semi-Thue systems. This problem can be reduced to the partial confluence problem for strict, finite semi-Thue systems by sending every instance (S, f) of the first problem to an instance $(\psi(S), \psi(f))$ of the second problem where ψ is a suitable homomorphism (such a trick has been described in the proof of Proposition 5.18). Proposition 5.32 follows from this reduction and Proposition 4.7. □

5.35. Remark. Theorem 5.27 has been proved independently in [30] in the particular case where S is a monadic (instead of basic) and R is reduced to a single word (instead of being any rational set). Proposition 5.32 is also proved in [30] (by a reduction different from ours).

6. Conclusion

Let us summarise in two tables the decidability results which are known about the five problems considered in Section 4 for finite semi-Thue systems, regular semi-Thue systems and finite c-systems (Table 1) and the results which are known about the three problems considered in Theorem 5.17 for finite semi-Thue systems (Table 2). In these tables, "yes", "no", "eq", "?" mean, respectively, that the problem is decidable, undecidable, recursively equivalent to the equivalence problem for dpda or that the decidability of the problem is an open question. An arrow from

Table 1

CLASS of rewriting systems / PROBLEM	Finite S.T.S. basic	Finite S.T.S. left basic	Finite S.T.S. no restriction	Regular S.T.S. basic	Regular S.T.S. left basic	Regular S.T.S. no restriction	Finite c-systems basic	Finite c-systems left basic	Finite c-systems no restriction
Confluence problem (for strict systems)	yes	yes	yes [26]	yes	yes	no [28]	yes	yes (Prop. 4.2)	no [28] (Prop. 4.1)
Refinement problem (for confluent, strict systems)	yes	yes	yes (obvious)	yes	yes	?	yes	yes (Prop. 4.3)	no
Equivalence problem (for confluent, strict systems)	yes	yes	yes	yes	yes	yes [24]	yes	yes	no (Prop. 4.5)
Class inclusion problem (for confluent, strict systems)	no [34]	no	no	no	no	no	no	no	no
Class equivalence problem (for confluent, strict systems)	yes [34]	? eq (Th. 5.17)	no (Prop. 4.7)	yes [35]	? eq	no	? eq (Th. 3.1)	? eq (Th 3.1)	no

Table 2

Problem	Additional restriction on semi-Thue systems: Basic	Left-basic	No restriction
$w_1 \equiv_{[f]_{\overset{*}{\leftrightarrow}_S}} w_2$	yes	eq	no
S confluent, strict, finite	(stated in [34, p. 310])	(Theorem 5.17)	(Proposition 4.7 + Lemma 5.26)
$[f]_{\overset{*}{\leftrightarrow}_{S_1}} = [f]_{\overset{*}{\leftrightarrow}_{S_2}}$	yes	eq	no
S_1, S_2 confluent, strict, finite	[34, Theorem III-2]	(Theorem 5.17)	(Proposition 4.7)
$\langle f\rangle_{\overset{*}{\leftrightarrow}_S} = [f]_{\overset{*}{\leftrightarrow}_S}$	yes	eq	no
S strict, finite	(Theorem 5.27)	(Theorem 5.17)	(Proposition 5.32)

a problem P toward a problem P' means that from the result concerning P, one can immediately deduce the result concerning P'.

In Table 1 it appears that, though c-systems are more general than regular semi-Thue systems, they have (up to our knowledge) similar decidability properties (with only one exception: the equivalence problem is decidable for confluent, strict, rational semi-Thue systems while it is not for confluent, strict, finite c-systems). This strengthens the intuition that the equivalence problem for dpda should be decidable.

Table 2 can be considered from two points of view. From the point of view of the equivalence problem for dpda:

(1) we have reduced the general problem to the equivalence problem for a strict subfamily of the family of dcfls: the subfamily of languages of the form $[f]_{\leftrightarrow_S}$ where f is a word and S is a left-basic, confluent, strict, finite, semi-Thue system;

(2) the three reductions given in Theorem 5.17 allow a new method for attacking the equivalence problem for dpda which consists of studying finitely generated congruences.

From the point of view of the decidability properties of finite semi-Thue systems we have classified the three problems involved in Table 2 in three classes:
- the decidable problems;
- the undecidable problems;
- the "very difficult" problems, this notion of "very difficult" being defined by the recursive equivalence with the equivalence problem for dpda.

Acknowledgment

This article corresponds to Chapters IV and V of the doctoral dissertation realised by the author at the University of Paris 7, under the direction of Luc Boasson and Jean-Michel Autebert [37].

References

[1] A.V. Aho and J.D. Ullman, *The Theory of Parsing, Translation and Compiling* (Prentice-Hall, Englewood Cliffs, NJ, 1972).

[2] C. Beeri, An improvement on Valiant's decision procedure for equivalence of deterministic finite-turn pushdown automata, *Theoret. Comput. Sci.* **3** (1976) 305–320.

[3] M. Benois and J. Sakarovitch, On the complexity of some extended word problems defined by cancellation rules, *Inform. Process. Lett.*, to appear.

[4] J. Berstel, Congruences plus que parfaites et langages algébriques, Séminaire d'informatique théorique de Paris 6, Paris 7, année 1976–1977.

[5] J. Berstel, *Transductions and Context-free Languages* (B.G. Teubner, Stuttgart, 1979).

[6] L. Boasson, Grammaires à non-terminaux séparés, in: *7th ICALP*, Lecture Notes in Computer Science, **85** (Springer, Berlin, 1980) 105–118.

[7] L. Boasson and M. Nivat, Centers of context-free languages, LITP technical report no. 84-44, 1984.

[8] L. Boasson and G. Sénizergues, NTS languages are deterministic and congruential, *J. Comput. System Sci.* **31** (1985) 332–342.

[9] R.V. Book, Thue systems as rewriting systems, *J. Symbolic Comput.* **3** (1987) 39-68.

[10] R.V. Book, M. Jantzen and C. Wrathall, Monadic Thue systems, *Theoret. Comput. Sci.* **19** (1982) 231-251.

[11] P. Butzbach, Une famille de congruences de Thue pour lesquelles l'équivalence est décidable, in: *1st ICALP* (North-Holland, Amsterdam, 1973) 3-12.

[12] P. Butzbach, Sur l'équivalence des grammaires simples, in: J.P. Crestin and M. Nivat, eds. *Actes des Premières Journées d'Informatique Théorique* (Bonascre, ENSTA, 1973).

[13] D. Caucal, Recritures suffixes de mots, Publication interne de l'IRISA, no. 413, June 1988.

[14] L. Chottin, Strict deterministic languages and controlled rewriting systems, in: *Proc. 6th ICALP*, Lecture Notes in Computer Science (Springer, Berlin, 1979).

[15] L. Chottin, Langages algébriques et systèmes de réécriture rationnels RAIRO, *Inform. Théor.* **16** (1982) 1-20.

[16] Y. Cochet, Sur l'algébricité des classes de certaines congruences définies sur le monoïde libre, Tech. Report Université de Rennes, 1971.

[17] Y. Cochet, Congruences quasi-parfaites et langages rationnels, Unpublished manuscript, 1975.

[18] Y. Cochet and M. Nivat, Une généralisation des ensembles de Dyck, *Israel J. Math.* **9** (1971).

[19] S. Eilenberg, *Automata, Languages and Machines* (Academic Press, New York/London, 1976).

[20] S. Greibach, A note on pushdown store automata and regular systems, *Proc. Amer. Math. Soc.* **18** (1967) 263-268.

[21] G. Huet, Confluent reductions: abstract properties and applications to term rewriting systems, *J. Assoc. Comput. Mach.* **27** (1980) 797-821.

[22] M. Jantzen, Thue congruences and complete string rewriting systems, Technical report, Universität of Hamburg, October 1985.

[23] F. Kierszenbaum, Les langages à opérateurs d'insertion, Thèse de 3 cycle, Université de Bordeaux I, 1979.

[24] P. Narendran, On the equivalence problem for regular Thue systems, *Theoret. Comput. Sci.* **44** (1986) 237-245.

[25] M. Nivat, On some families of languages related to the Dyck language, in: *2nd Annual Symp. on Theory of Computing* (ACM, 1970).

[26] M. Nivat and M. Benois, Congruences parfaites et quasi-parfaites, *Sém. Dubreil (Algèbre)* **25** (1972) 1-9.

[27] C. O'Dunlaing, Finite and infinite regular Thue systems, Ph.D. Dissertation, University of California at Santa Barbara, 1981.

[28] C. O'Dunlaing, Infinite regular Thue systems, *Theoret. Comput. Sci.* **2** (1983) 171-192.

[29] F. Otto, Some undecidability results for non-monadic Church-Rosser Thue systems, *Theoret. Comput. Sci.* **33** (1984) 261-278.

[30] F. Otto, On deciding the confluence of a finite string-rewriting system on a given congruence class, *J. Comput. System Sci.* **35** (1987) 285-310.

[31] M. Oyamaguchi, The equivalence problem for real-time d.p.d.a.'s, *J. Assoc. Comput. Mach.* **34** (1987) 731-760.

[32] J. Sakarovitch, Syntaxe des langages de Chomsky, essai sur le déterminisme, Thèse de doctorat d'état de l'université Paris VII, 1979.

[33] G. Sénizergues, Décidabilité de l'équivalence des grammaires NTS, Thèse de 3$^{\text{ième}}$ Cycle, Université de Paris VII, 1981.

[34] G. Sénizergues, The equivalence and inclusion problems for NTS languages, *J. Comput. System Sci.* **31** (1985) 303-331.

[35] G. Sénizergues, Church-Rosser controlled rewriting systems and equivalence problems for deterministic context-free languages, *Inform. and Comput.* **81** (3) (1989) 265-279.

[36] G. Sénizergues, A characterisation of deterministic context-free languages by means of right-congruences, *Theoret. Comput. Sci.* **70** (2) (1990) 213-232.

[37] G. Sénizergues, Sur la description des langages algébriques déterministes par des systèmes de réécriture confluents, Thèse d'Etat, Université Paris 7, Juillet 1987. Rapport LITP 88-39.

[38] L.G. Valiant, The equivalence problem for deterministic finite-turn pushdown automata, *Inform. and Control* **25** (1974) 123-133.